## Government Polytechnic ,Washim

A

MICRO PROJECT REPORT ON

**“MOVIE RECOMMENDATION SYSTEM ”**

****

**Submitted By group of**

**1. Pavan P. Faltankar ( 1700310215)**

**2. Sacchidanand B.Linge ( 1800310215)**

**3. Vaibhav S.Ingle ( 1800310207)**

**4. Pavan S.Ingole ( 1700310216)**

***(Final Year Diploma in Information Technology)***

**Principal**

**Dr. V.R.Mankar**

***( Government Polytechnic, Washim)***

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**2019- 2020**

**Government Polytechnic,Washim CERTIFICATE**

**This is to certify that**

**1. Pavan P. Faltankar**

**2. Sacchidanand B.Linge**

**3. Vaibhav S.Ingle**

**4. Pavan S.Ingole**

**Final year Students of Information Technology have submitted a project report on**

**“MOVIE RECOMMENDATION SYSTEM ”**

**During the sixth semester of the academic year 2019-2020 in a satisfactory manner in the partial fulfillment for the requirement for the Diploma in**

**“Information Technology” awarded by**

**Maharashtra State Board of Technical Education, Mumbai.**

**Mrs. M.R.Lolure** **Mrs M. R. Lolure *( Lecturer in Info. Tech. Department)*** *(I/C HOD ,Information Technology Department)*

**Dr. V.R.Mankar**

**Principal**

***( Government Polytechnic, Washim)***

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Contents** | **Page no.** |
| 1 | Abstract | 2 |
| 2 | Introduction | 3 |
| 3 | Related Work | 7 |
| 4 | Research Methodology | 8 |
| 5 | Data Description | 10 |
| 6 | Proposed Algorithm | 11 |
| 7 | Overcome The Problems | 13 |
| 8 | Conclusion | 14 |

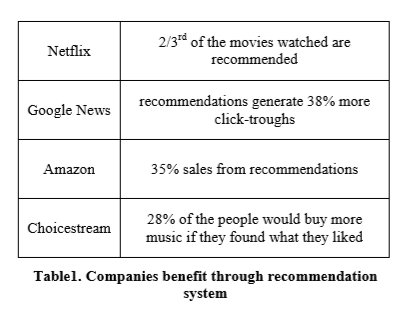
**ABSTRACT**

Now a day’s recommendation system has changed the style of searching the things of our interest. This is information filtering approach that is used to predict the preference of that user. The most popular areas where recommender system is applied are books, news, articles, music, videos, movies etc. In this paper we have proposed a movie recommendation system named MOVREC. It is based on collaborative filtering approach that makes use of the information provided by users, analyzes them and then recommends the movies that is best suited to the user at that time. The recommended movie list is sorted according to the ratings given to these movies by previous users and it uses K-means algorithm for this purpose. 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. This system has been developed in PHP using Dreamweaver 6.0 and Apache Server 2.0. The presented recommender system generates recommendations using various types of knowledge and data about users, the available items, and previous transactions stored in customized databases. The user can then browse the recommendations easily and find a movie of their choice ABSTRACT Now a day’s recommendation system has changed the style of searching the things of our interest. This is information filtering approach that is used to predict the preference of that user. The most popular areas where recommender system is applied are books, news, articles, music, videos, movies etc. In this paper we have proposed a movie recommendation system named MOVREC.

**INTRODUCTION**

In today’s world where internet has become an important part of human life, users often face the problem of too much choice. Right from looking for a motel to looking for good investment options, there is too much information available. To help the users cope with this information explosion, companies have deployed recommendation systems to guide their users. The research in the area of recommendation systems has been going on for several decades now, but the interest still remains high because of the abundance of practical applications and the problem rich domain. A number of such online recommendation systems implemented and used are the recommendation system for books at Amazon.com , for movies at MovieLens.org, CDs at CDNow.com .

Recommender Systems have added to the economy of the some of the e-commerce websites (like Amazon.com) and Netflix which have made these systems a salient parts of their websites. A glimpse of the profit of some websites. s is shown in table below:



Recommender Systems generate recommendations; the user may accept them according to their choice and may also provide, immediately or at a next stage, an implicit or explicit feedback. The actions of the users and their feedbacks can be stored in the recommender database and may be used for generating new recommendations in the next user-system interactions. The economic potential of theses recommender systems have led some of the biggest e-commerce websites (like Amazon.com, snapdeal.com) and the online movie rental company Netflix to make these systems a salient part of their websites. High quality personalized recommendations add another dimension to user experience. The web personalized recommendation systems are recently applied to provide different types of customized information to their respective users. These systems can be applied in various types of applications and are very common now a day.

We can classify the recommender systems in two broad categories:

1. **Collaborative filtering approach:-**

2. **Content-based filtering approach:-**

**1.** **Collaborative filtering Collaborative filtering** system recommends items based on similarity measures between users and/or items. The system recommends those items that are preferred by similar kind of users. Collaborative filtering has many advantages.

1. It is content-independent i.e. it relies on connections only.
2. Since in CF people makes explicit ratings so real quality assessment of items are done.
3. It provides serendipitous recommendations because recommendations are base on user’s similarity rather than item’s similarity.

**2. Content-based filtering Content-based filtering** is based on the profile of the user’s preference and the item’s description. In CBF to describe items we use keywords apart from user’s profile to indicate user’s preferred liked or dislikes. In other words CBF algorithms recommend those items or similar to those items that were liked in the past. It examines previously rated items and recommends best matching item. There are various approaches proposed in various research papers listed below.

These approaches are often combined in Hybrid Recommender Systems. An earlier study all for the recommendation of movies through MOVIEGEN had certain drawbacks such as , it asks a series of questions to users which was time taking .

On the other hand it was not user friendly for the fact that it proved to be stressful to a certain extent. Keeping in mind these shortcomings, we have developed Movie REC, a movie recommendation system that recommends movies to users based on the information provided by the users themselves.

In the present study, a user is given the option to select his choices from a set of attributes which include actor, director, genre, year and rating etc. We predict the users choices based on the choices of the previous visited history of users.

**RELATED WORK**

Many recommendation systems have been developed over the past decades. These systems use different approaches like collaborative approach, content based approach, a utility base approach, hybrid approach etc. Looking at the purchase behavior and history of the shoppers, Lawrence et al. 2001 presented a recommender system which suggests the new product in the market. To refine the recommendation collaborative and content based filtering approach were used. To find the potential customers most of the recommendation systems today use ratings given by previous users. These ratings are further used to predict and recommend the item of one’s choice. In 2007 Weng, Lin and Chen performed an evaluation study which says using multidimensional analysis and additional customer’s profile increases the recommendation quality. Weng used MD recommendation model (multidimensional recommendation model) for this purpose. multidimensional recommendation model was proposed by Tuzhilin and Adomavicius (2001).

**RESEARCH METHODOLOGY**

**The Basic K-means Algorithm:-**

The original K-means algorithm was proposed by MacQueen [20] .The ISODATA algorithm by Ball and Hall[22] was an early but sophisticated version of k-means. Clustering divides the objects into meaningful groups. Clustering is unsupervised learning. Document clustering is automatic document organization. In K-means clustering technique we choose K initial centroids, where K is the desired number of clusters. Each point is then assigned to the cluster with nearest mean i.e. the centroid of the cluster. Then we update the centroid of each cluster based on the points that are assigned to the cluster. We repeat the process until there is no change in the cluster center (centroid). Finally, this algorithm aims at minimizing an objective function, in this case a squared error function.

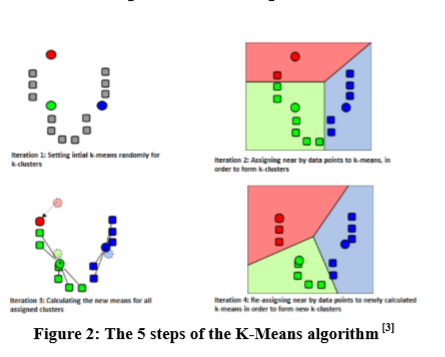
**The algorithm is composed of the following steps:**

1. Select K points as initial centroids.  
 2. Repeat.

3. From k clusters by assigning each point to its closest centroid.

4. Re-compute the centroid of each cluster.

5. Until Centroid do not change.



**DATA DESCRIPTION**

In proposed model we use a pre filter before applying Kmeans algorithm. The attributes used to calculate distance of each point from centroid are

1. Genre 2. Actor .

Different attributes have different weights. In our research we have found that the most appropriate recommendations that can be generated should be based on the ratings given to the movies by previous users, therefore we have given more importance to the rating attribute than other attributes. These ratings have been taken from www.imdb.com because perhaps it has the largest collection of movies along with the rating given to these movies by a large number of different users from different parts of the world. Another important parameter in our proposed model is total number of votes received by a particular movie. We have divided number of votes in to three categories that is less than or equal to 1000, more than 1000 but less than or equal to 10,000 and greater than 10,000.

**PROPOSED ALGORITHM**

Input: a number of movies: m

Output: a number of clusters: K

Step 1 : Select n movies from m movies n<m.

Step 2 : If n>20 then select top 20 movies from n movies based on ratings. Else display the output movies sorted by rating.

Step 3: If rating of movies x, y are equal

i.e. If Rx= Ry Then select those movies which have greater number of user votes.

Step 4: Assume K=4.

Step 5 :REPEAT (6, 7)

Step 6 : Chose initial centroid C1, C2, C3, C4.

Step 7 : Calculate Euclidean distance of all data points.

w.r.t. C1, C2, C3, C4 and re-compute the centroid of each cluster.

Step 8: UNTILL centroid does not change.

**CAHALLENGES FACED**

In developing any system the biggest challenge is to satisfy the end users for which the system is being developed. We also faced certain challenges while developing our system. Some of them are:-

* To have a system that is user friendly and easy to understand and use.
* To create a data set that has all relevant information about a particular movie.
* The biggest challenge was to have the most appropriate movie recommended list.

* To make our system diversifiable so that it can satisfy users of different geographical locations.

* To give weights to different attributes.

**OVERCOME THE PROBLEMS**

* The proposed system has been tested over a small group of people, and we have received a positive response from them. We have kept our system simple and interactive for this we have choose php and java script.

* For collecting information we have intensively search free online movie data bases and extract the information which was useful for our proposed system.

* To accurately recommend movie to user we have applied K-means clustering algorithm along with a pre filter.

* We have included movies in our database irrespective of their language or location so that users from all across the globe can use our system.
* For assigning weights to attributes and for giving priority to them we have conducted a survey on a group of people and on the basis of the result obtained we have prioritize our attributes.

**CONCLUSION**

In this project we have introduced MovieREC, a recommender system for movie recommendation. It allows a user to select his choices from a given set of attributes and then recommend him a movie list based on the cumulative weight of different attributes and using K-means algorithm. By the nature of our system, it is not an easy task to evaluate the performance since there is no right or wrong recommendation; it is just a matter of opinions. Based on informal evaluations that we carried out over a small set of users we got a positive response from them. We would like to have a larger data set that will enable more meaningful results using our system.