Synopsis of Project work on,

"TeenHub"

Submitted in partial fulfilment of the requirements

of the degree of

(Bachelor of Engineering)

by

Pranav Nilesh Jain

Kedar Kishor Kale

Paresh Raju Aher

Ganesh Dhanaram Choudhary

Supervisor:

(Deepali Maste)

(Computer Engineering)

AET's

Atharva College of Engineering

(2017-18)

AET'S
ATHARVA COLLEGE OF ENGINEERING

CERTIFICATE

This is to certify that

Pranav Nilesh Jain

Kedar Kishor Kale

Paresh Raju Aher

Ganesh Dhanaram Choudhary

Have satisfactorily completed the requirements of the Synopsis

On

"TeenHub"

As prescribed by the University of Mumbai for academic year 2017-18
Under the guidance of

Prof. Deepali Maste
Guide

Prof Mahendra Patil
Head (Computer Engineering)

Dr S.P.Kallurkar
Principal

Project Coordinator

Internal Examiner

External Examiner

College Seal
532

Index

List of Figures

Sr No	Figures	Page No
1	Context Level diagram	19
2	DFD Diagram	20
3	Sequence Diagram	23
4	E-R Diagram	24
5	Control Flow Diagram	25
6	Gantt Chart	27

List of Tables

Sr No	Tables	Page No
1	Gantt Chart Week Duration Table	27

Abstract

The domain of our project falls into two categories: Machine Learning and Advanced Web Application. In this project we are planning to visualize the Machine Learning algorithms on the web applications by using smart data visualization tools, algorithms and techniques and not limiting them to just run onto terminals. It will be an all in one package for entertainment news and ratings.

Our project is an entertainment hub for people of all age groups. It is divided into four main parts:

1. Movies Section
2. Songs Section
3. News Section
4. Games Section

The users will have one common login for all the 4 parts. The users will get a dashboard to see their own data and activities on interactive charts. The user can login as a guest or make an account for self. The recommendation engine will recommend the user movies based on his previous ratings for movies . Also , when the user browses through movies , songs or games , KNN algorithm[3] will make sure to display the user with suggestions similar to the current item. We have also included a news app to keep the user updated with the latest of news in the entertainment industry.

The project will have a custom search engine based on TF-IDF[4] algorithm to navigate through the application . The movie recommendation engine will be based on KNN algorithm and will be trained on a dataset of not less than 1,00,000 real ratings.

1. Introduction

1.1 Need

On the Internet, where the number of choices is overwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users. Recommender systems solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services.

The explosive growth in the amount of available digital information and the number of visitors to the Internet have created a potential challenge of information overload which hinders timely access to items of interest on the Internet[2]. Information retrieval systems, such as Google, and Altavista have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for recommender systems more than ever before.

1.2 Basic Concept

TeenHub is an advanced Web Application which uses Machine Learning Algorithms. It is a recommendation engine for movies, songs, news and games. The project also houses a custom search engine which makes navigation a lot easier.

It uses TF-IDF[4] algorithm for the search engine and KNN algorithm for the recommendation system of the application. TeenHub promises to show only relevant data to the users with the help of these algorithms.

To make sure that the models which will be created for each user, remain accurate TeenHub will rely on more than 1,00,000 real ratings made by humans and not randomized ratings. This also makes sure that the overall accuracy of the system also remains high.

1.3 Application

- Teenhub has various entertainment fields such as movies, songs, games and news which help to user to get all kind of information on the same platform.
- It Uses user previous activity such as rating, history to recommend new movies, games, or songs and it makes user-friendly environment.
- Teenhub make easy for the user to access all field just by single step registration.
- In order to select best games or movie, it also provide weekly, monthly and yearly ranking of games and songs, based on user rating in particular genre.
- Teenhub also include news related to games and movies and weekly popular movie to help user for getting best choice.
- It allows user to rate movie, review, rate game in any appropriate genre and accordingly system recommend movies, games or songs to him/her.

2. Review of Literature

Teenhub is a recommendation system, which recommends movie, games and songs based on rating entered by different users. This can be used at time of deciding whether to watch particular movie or not, or whether to buy game or not.

Recommendation system is mainly based on user based collaborative filtering and item based collaborative filtering. In user based collaborative filtering similarity between users is evaluated using their rating .In item based collaborative filtering similarity between items is evaluated. Since choice of users vary frequently[5], so here we are using item based collaborative filtering.

TF-IDF (Term Frequency Inverse Document Frequency) algorithm[4] is used for search engine for providing relevant result of movie, game and song. TF-IDF weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus.

K-nearest neighbor algorithm (KNN)[3] is used for classification of movie, games and songs. For example all racing games are grouped together. KNN algorithm will make sure to display the user with suggestions similar to the current item.

Here we are using these algorithms in single web based application.

3. Report on the Existing System

Current systems employ user based collaborative filtering which has its fair share of disadvantages.

User based CF recommend items by finding similar users. This is often harder to scale because of the dynamic nature of users.

Data sparsity is seen as a key disadvantage of user-based CF[5]. It is often assumed that data sparsity may cause small number of co-rated items or no such ones between two users, resulting in unreliable or unavailable similarity information, and further incurring poor recommendation quality. Similarity estimates between items are more likely to converge over time than similarities between users.

We can compute and cache similarities that converge, which can give item based recommenders a performance advantage

4. Aim and Objectives

The aim of this project is to provide a one place solution for all entertainment needs of users of all age group.

The objectives are:

1. To implement recommendation engine for movies, songs, games and news.

2. To implement a custom search engine.

3. To design a responsive design for the web application.

4. To implement a dashboard for the users to view their activities.

5. Problem statement

Nowadays, it become very hard to find the information related entertainment activity on same platform and its very time taking to gather information regarding movies, games, news or songs though, Due to diversity in platform, user couldn't focus on other activity such as news, ranking and it makes hard for user to get complete information in his favorites genre.

6. Proposed System for Project

The proposed system has three main components:

- Similarities among items : In this step, correlation between selected item rating and all other items ratings in database is evaluated[2]. Item with highest correlation value is recommended to user.

- Classification : This step applied at start of application. In this step classification of movies, games and songs based on different genre. For each possible genre there is panel which display movies, games and songs related to genre.

- Search Engine : This step consider tf-idf weight, which is composed by two terms: the first computes the normalized Term Frequency (TF), aka. The number of times a word appears in a document, divided by the total number of words in that document; the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.
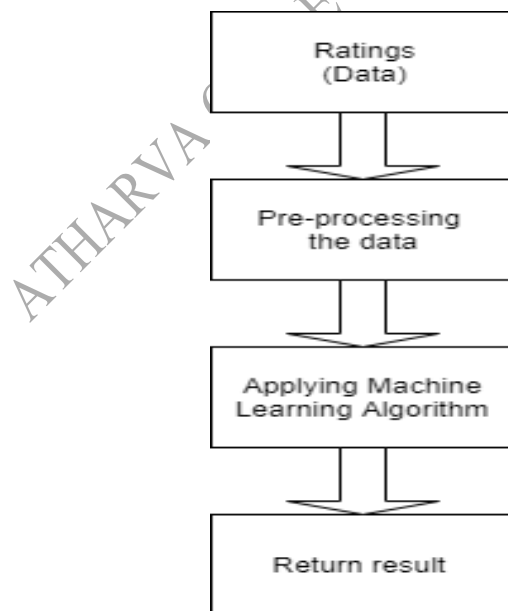

Fig 6.1

7. Requirement Analysis (SRS)

Functional requirements:

Software requirements

- Client side : Any OS supporting any browser
- Server side : django framework, MySQL, Python, NumPy, Pandas, SciPy

Hardware requirements

- Client side : Any device (mobile or desktop) with decent hardware as it is a web application and requires less computation power on the client side
- Server side : Minimum quad-core processor with 4gb of ram (requirement increases as server traffic increases)

Non-Functional requirements

- Performance requirements - System is designed using numpy and pandas libraries which provide optimum performance. Also because the application is a webapp it can work on any operating system with a browser.
- security requirements - The system provides authentication to users with the help of a login system.

8. Scope

The current scope of the project is to cater all the entertainment focused things (like movies, songs, games and news) to users. The plan is to create new models by tweaking settings and using new algorithms, so that the application can return or predict accurate results. Also, in the design aspect of the application, the focus is on to implement mutative design for the users.

Also, special care is taken to make the structure of the application as modular as possible. This makes it much easier to add new features in future and it also reduces inter-dependence between the modules.

Each user will be given a dashboard where that particular user's data will be shown in interactive charts with the help of javascript libraries. The users can view, edit their activities from this dashboard.

Future modifications are to include advanced features like predicting movie ratings which are not released based on many factors like, director, cast ratings, budget, etc. Also, future plan is to build native applications for the mobile phones.

9. Design Details

The design includes:

a) Context Level diagram:

Context Level Diagram, in engineering is a diagram that defines the boundary between the system, or part of a system, and its environment, showing the entities that interact with it.

b) DFD Diagram:

Data Flow Diagram is a graphical representation of the "flow" of data through an information system, modelling its process aspects.

c) Sequence Diagram:

Sequential Diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart.

d) E-R Diagram:

Entity Relationship Diagram shows the relationships of entity sets stored in a database.

e) Control Flow Diagram:

This diagram shows the flow of control through the system. The flow starts with the user interacting with the system in search for recommendation.
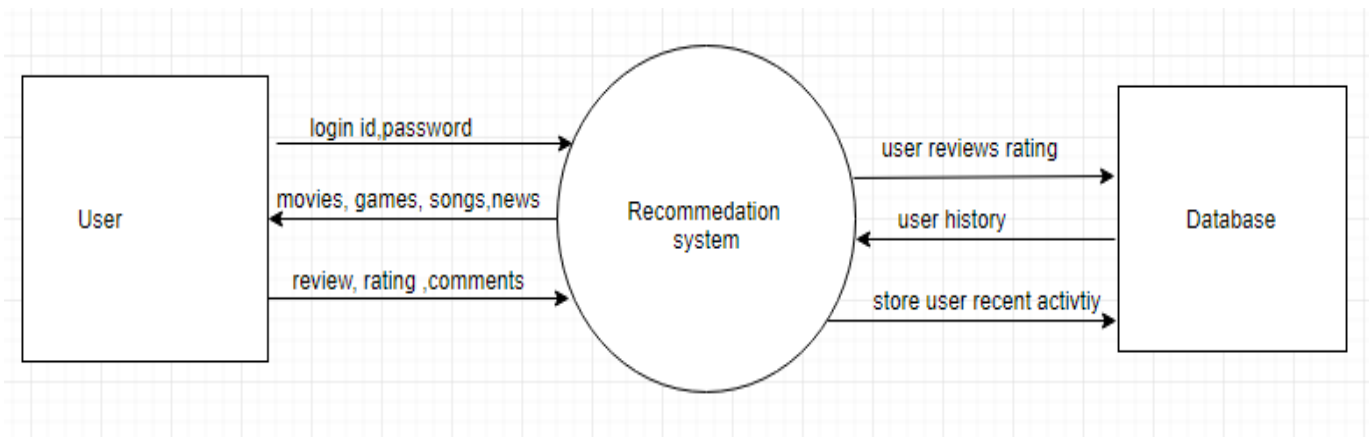
9.1 context level diagram

Fig 9.1

Context Level Diagram, in engineering is a diagram that defines the boundary between the system, or part of a system, and its environment, showing the entities that interact with it. The recommendation system takes user selection and ratings as input and provides recommendations in return.
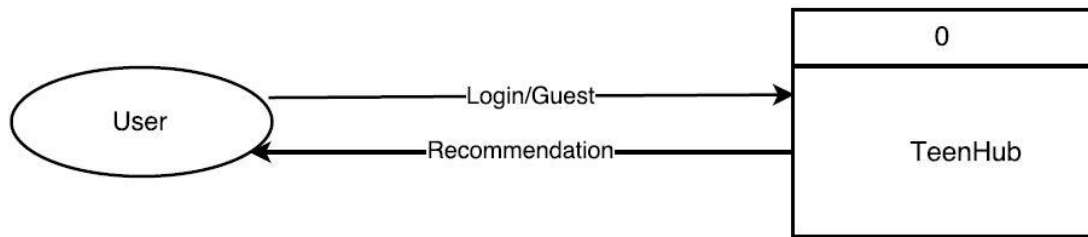
9.2 DFD Diagram

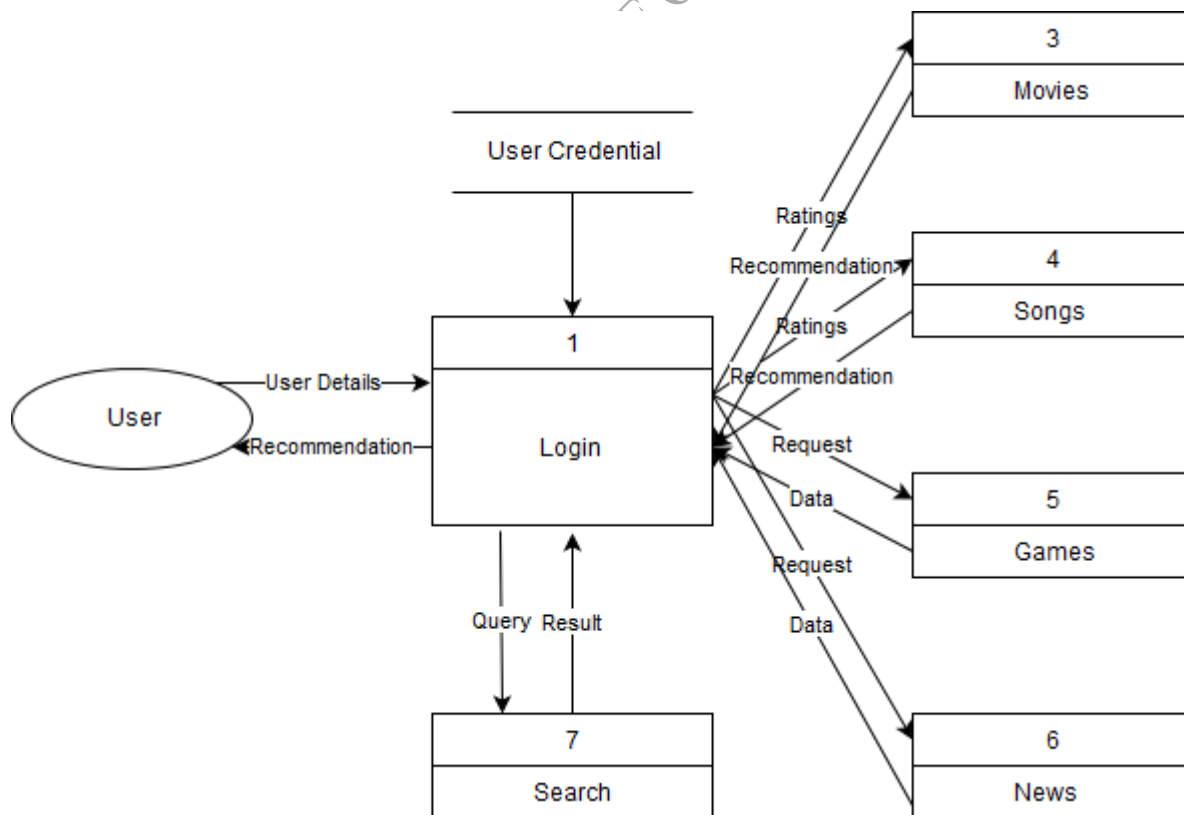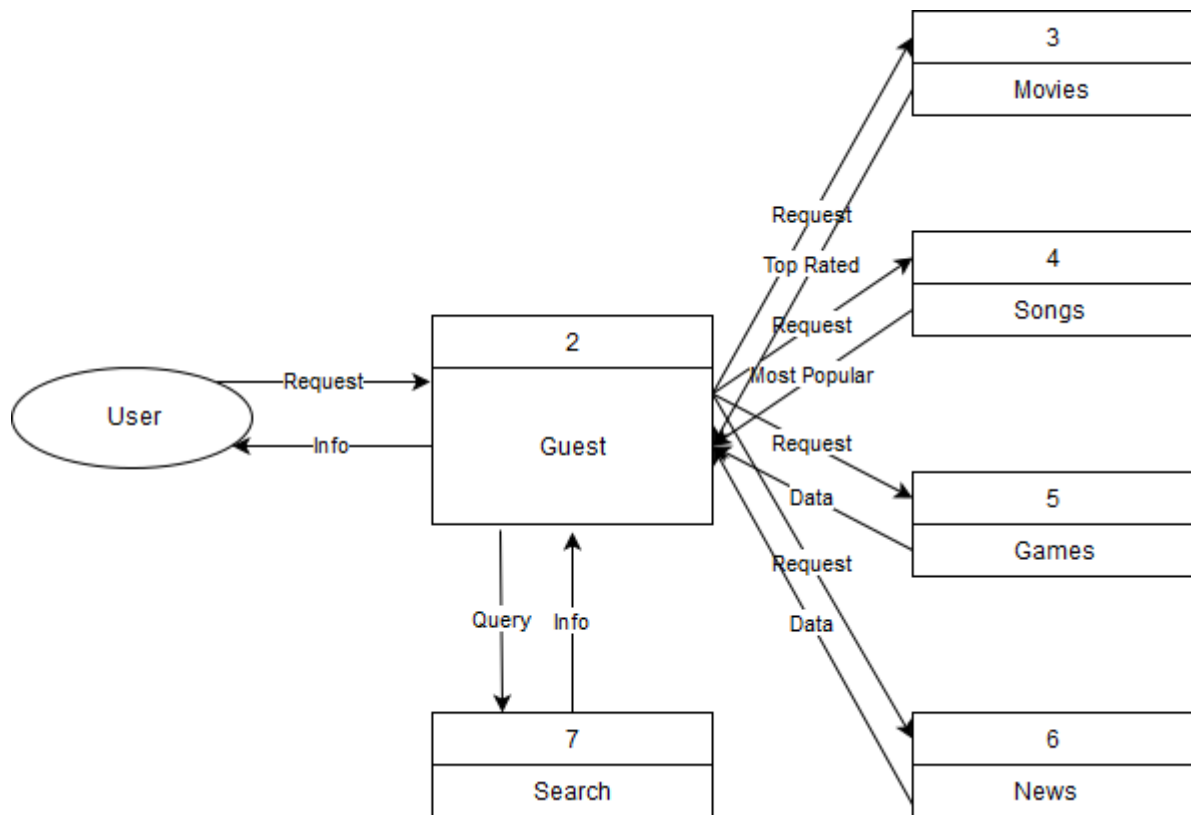Level 0 DFD:

Fig 9.2.1

Level 1 DFD

Fig 9.2.2

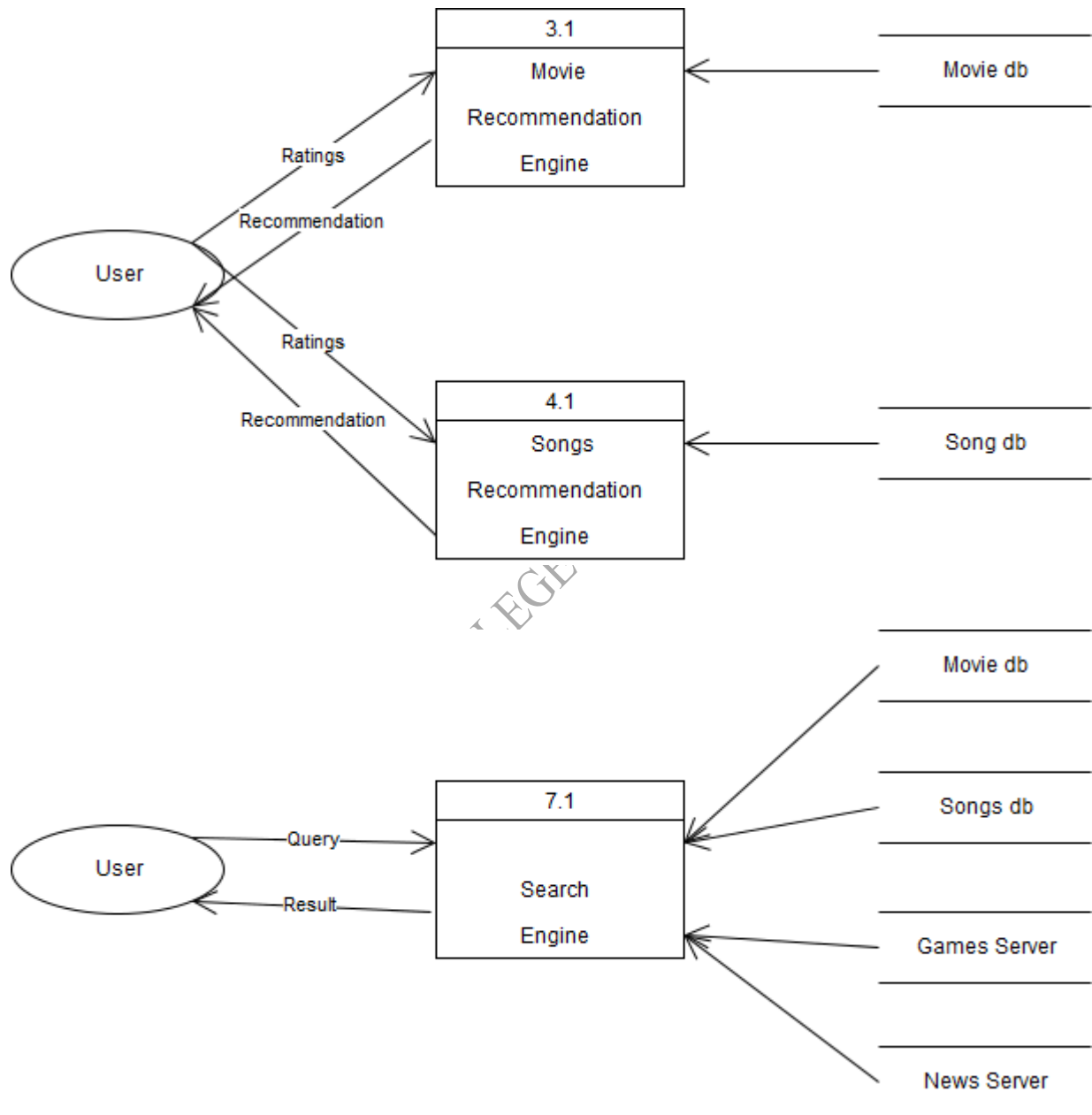*Fig 9.2.3*

Level 2 DFD:

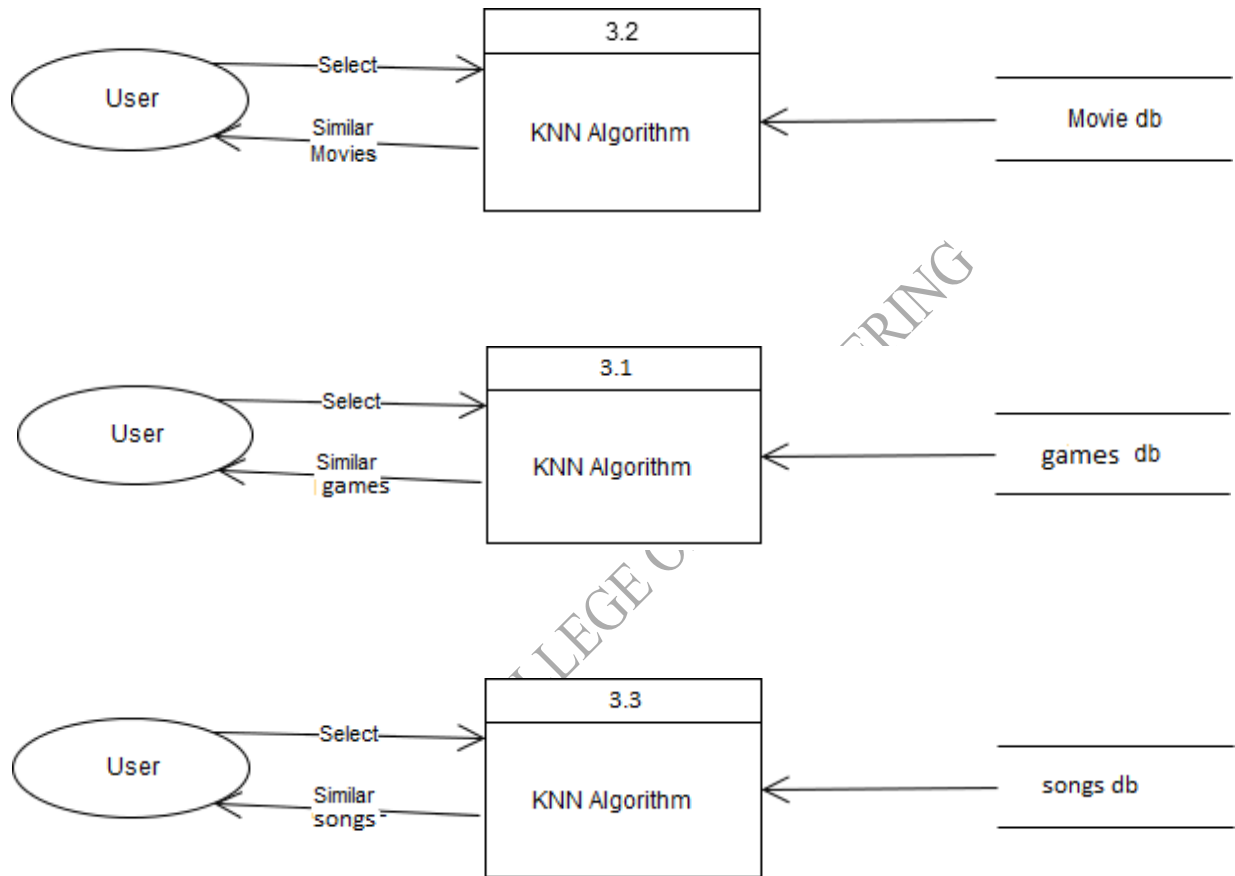*Fig 9.2.4*

Level 3 DFD:

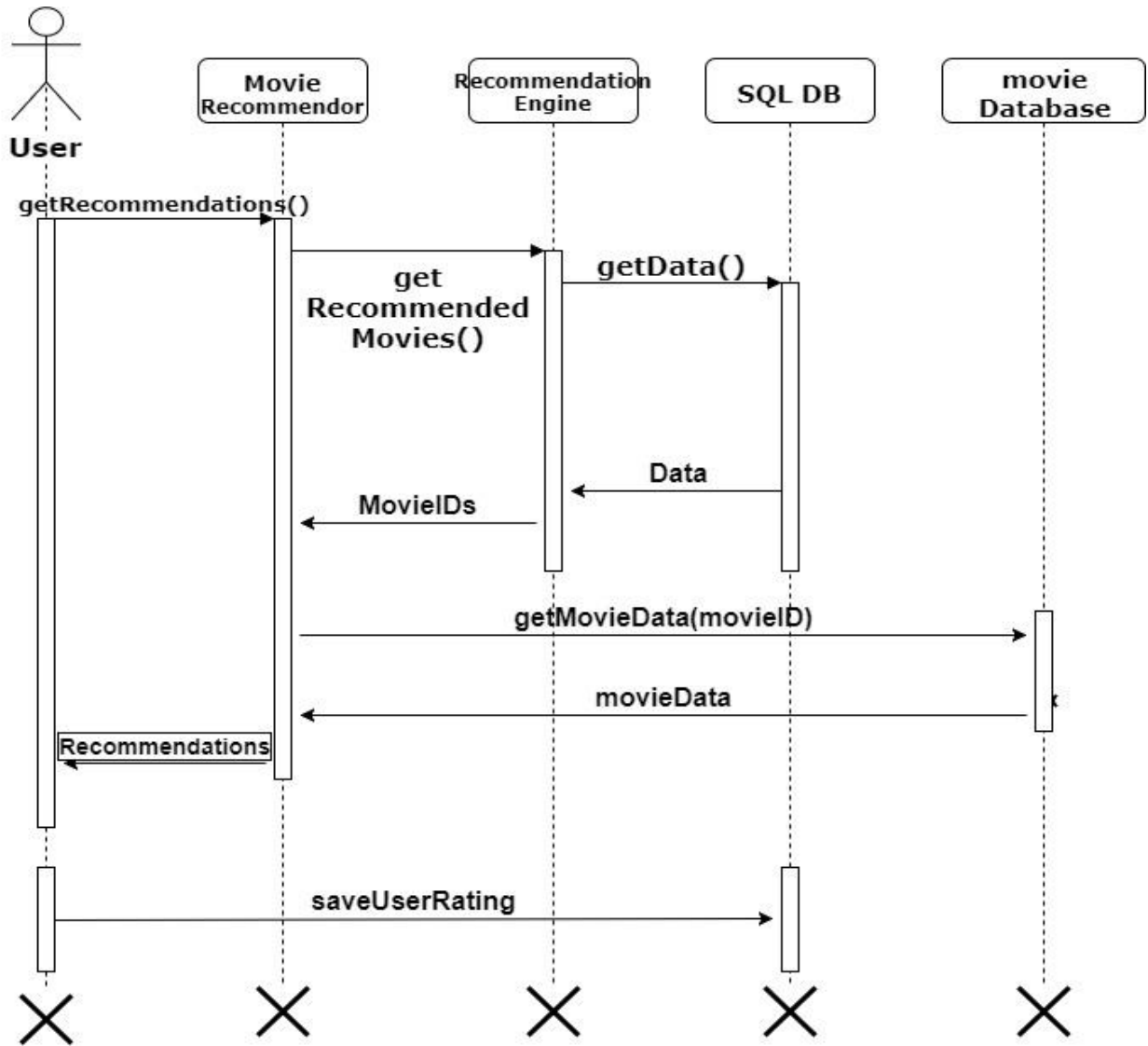*Fig 9.2.5*

9.3 Sequence Diagram

Fig 9.3.1

The above figure 9.3.1 shows the sequence in which the user requests are handled and the recommendations are generated.
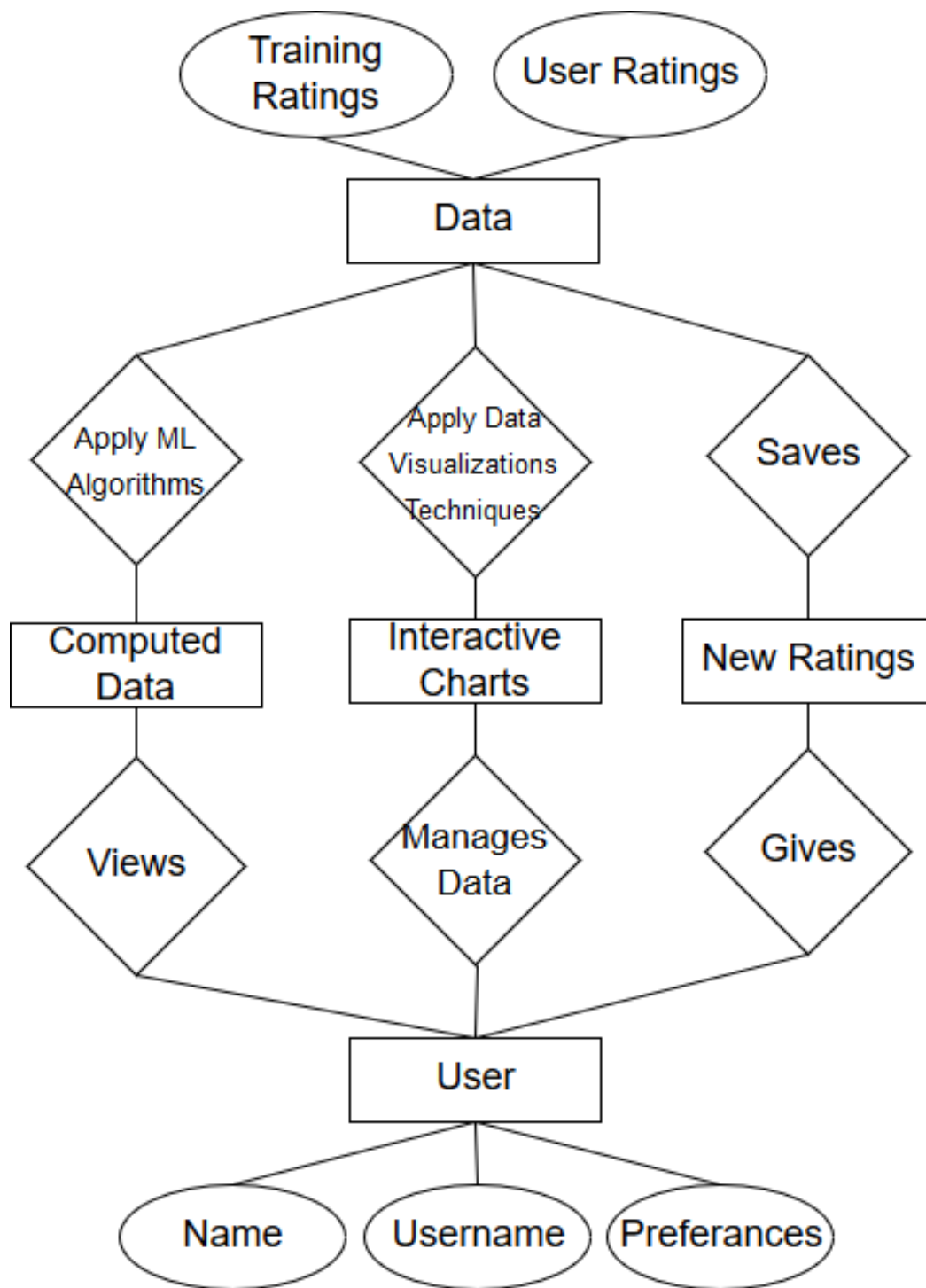
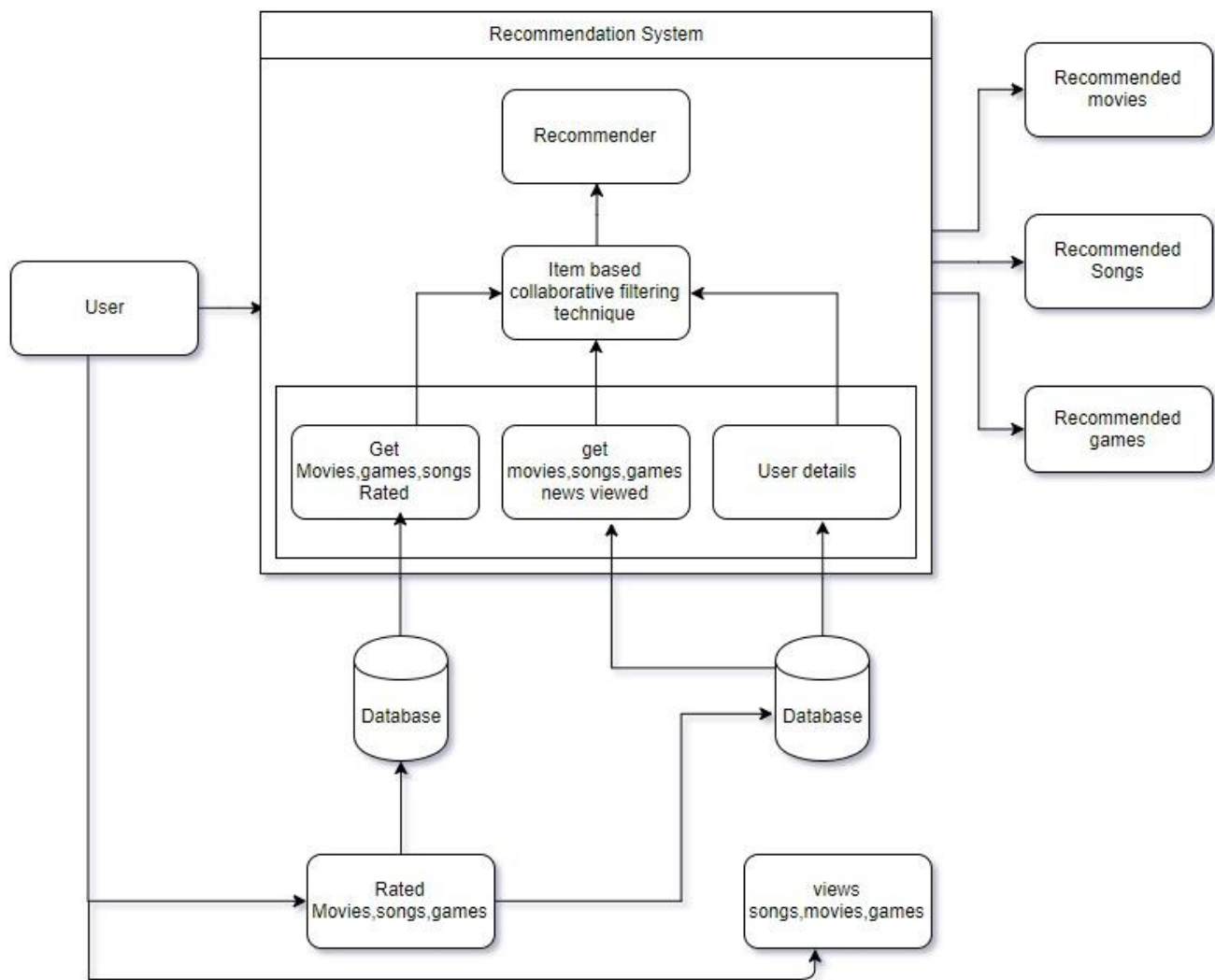9.4 ER Diagram

Fig 9.4.1

9.5 Control flow diagram

Fig 9.5.1

This diagram shows the flow of control through the system. The flow starts with the user interacting with the system in search for recommendation. The control flows from the user to the recommendation engine to the database and back again.

10. Implementation plan

10.1 Software and Hardware requirements :

Software requirement

1. Operating System: Windows 7/8/10

2. Library: Chart.js, numpy, pandas, scipy

3. Software: XAMPP, Canopy

4. Programming language: python, javascript, html and css

Hardware requirement

1. Processor: 1.6 GHz processor

2. RAM:4GB RAM

3. Memory: 300MB

10.2 Gantt chart

Month	July				August			
week	1	2	3	4	5	6	7	8
Project concept	▓	▓						
Analysis on technologies used			▓	▓				
Project Architecture					▓	▓	▓	
Abstract								▓
Presentation								
Literature Survey								
Synopsis								

Month	September				October		
week	9	10	11	12	13	14	15
Project concept							
Analysis on technologies used							
Project Architecture							
Abstract							
Presentation	▓	▓					
Literature Survey			▓	▓			
Synopsis						▓	▓

Fig 10.2.1

11. Methodology

a) Initiation (Requirement Analysis):

To gather data to develop a recommendation system for movies, songs, games and news. Then the next process to perform the preprocessing procedure. Preprocessing is the process in which data is prepared for the machine learning algorithms. Any incomplete is either removed or edited with the help preprocessing algorithms which are available in sklearn.

b) Feasibility Analysis:

The project is a web application which just requires a browser for it to render and do the computational tasks. Therefore, no downloading of software is required at the client side. However, a constant internet connection is required and the speed is recommended to be 2mbps+. The project uses the traditional html, css and javascript for the frontend and user side computational work. The backend server is of Django and uses jinja template. The core machine learning algorithms are written in python.

c) Design:

The design is a very important aspect of any web application and this project is no exception. Here, we are trying to implement mutative design which directly targets and caters the needs of the users. This project also follows a mobile-first design approach which makes sure that application smoothly on mobile devices.

d) Operation and Maintenance:

The developers of this project do not need to stop the functionality of the application to fix a particular part or for maintenance work, all thanks to the modular structure of the web application. The backend part is not visible to the users and thus implementing or inserting new features is very easy since the complexities are hidden behind the clever and user-friendly frontend design.

12. Conclusion

The users will have one common login for all the 4 parts. The users will get a dashboard to see their own data and activities on interactive charts. The user can login as a guest or make an account for self. The recommendation engine will recommend the user movies based on his previous ratings for movies. Also, when the user browses through movies, songs or games, KNN algorithm will make sure to display the user with suggestions similar to the current item. We have also included a news app to keep the user updated with the latest of news in the entertainment industry.

The project will have a custom search engine based on TF-IDF algorithm to navigate through the application. The movie recommendation engine will be based on KNN algorithm and will trained on a dataset of not less than 1,00,000 real ratings

Although, Teenhub provide facility for user to get various entertainment information on single platform and due to recommendation based system it helps user to get right content and it also provide ranking of movies and games which help user to select proper movie or game and it has daily news feature which keep user updated.

13. Acknowledgement

14. Literature Cited

[1] Dietmar Jannach and Simon Fischer, Recommendation-based Modeling Support for Data Mining Processes, Germany and Simon Fischer Rapid Miner GmbH, Germany, Proceedings of the 8th ACM Conference on Recommender system, pages 337-340, October 2014.

[2] Ali Heydarzadegan, Yaser Nemati, Mohsen Moradi, Evaluation of Machine Learning Algorithms in Artificial Intelligence, IJCSMC, Vol. 4, Issue. 5, May 2015, pg.278 – 286

[3] Tian Xia, Yanmei Chai, An Improvement to TF-IDF: Term Distribution based Term Weight Algorithm, JOURNAL OF SOFTWARE, VOL. 6, NO. 3, MARCH 2011

[4] Ramos, Juan. (2003). Using TF-IDF to determine word relevance in document queries.

[5] G. Linden ; B. Smith ; J. York, Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Computing (Volume: 7, Issue: 1, Jan/Feb 2003), pages 76-80

15. Appendix

1. KNN Algorithm:

   K-Nearest Neighbors Algorithm is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space.

2. TF-IDF Algorithm:

   Term Frequency–Inverse Document Frequency Algorithm is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval, text mining, and user modeling.

3. ML: Machine Learning

   Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.

4. DFD Diagram:

   Data Flow Diagram is a graphical representation of the "flow" of data through an information system, modelling its process aspects.

5. Context Level Diagram:

   Context Level Diagram, in engineering is a diagram that defines the boundary between the system, or part of a system, and its environment, showing the entities that interact with it.

6. Sequential Diagram:

   Sequential Diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart.

7. Gantt Chart:

Gantt Chart is a type of bar chart that illustrates a project schedule. Gantt charts illustrate the start and finish dates of the terminal elements and summary elements of a project.

8. E-R Diagram:

Entity Relationship Diagram shows the relationships of entity sets stored in a database.

9. Mutative Design:

Mutative Design is a user centered design in which the users are given the priority before designing any particular part of the project.