MovieBuzz

Submitted in partial fulfilment of requirements

For the degree of

Bachelor in Information Technology

by

Sanjana Anaokar Roll No: 1214022

Ria Marwaha Roll No: 1214023

Pooja Karkera Roll No: 1214047

Chaitali Tanawade Roll No: 1214048

Prof. Deepti Patole

Department of Information Technology

K.J.Somaiya College of Engineering, Mumbai-400077

(Autonomous College Affiliated to University of Mumbai)

Batch 2012-2016

K.J.Somaiya College of Engineering, Mumbai-400077

(Autonomous College Affiliated to University of Mumbai)

Certificate

This is to certify that the Dissertation entitled "MovieBuzz" is bona fide record of the dissertation work done by

1. Sanjana Anaokar (1214022)
2. Ria Marwaha (1214023)
3. Pooja Karkera (1214047)
4. Chaitali Tanawade (1214048)

in the year 2015-16 under the guidance of Prof. Deepti Patole of Department of Information Technology in partial fulfillment of requirement for the Bachelors degree in Information Technology of University of Mumbai.

Guide/Co-Guide Head Of Department

Principal

Date :

Place : Mumbai - 400077

K.J.Somaiya College of Engineering, Mumbai-400077

(Autonomous College Affiliated to University of Mumbai)

Certificate of Approval of Examiners

We certify that this project report entitled "MovieBuzz" is bona fide record of project work done by

1. Sanjana Anaokar(1214022)
2. Ria Marwaha(1214023)
3. Pooja Karkera(1214047)
4. Chaitali Tanawade(1214048)

This project is approved for the award of Bachelors Degree in Information Technology of University of Mumbai.

External Examiner Internal Examiner

Date :

Place : Mumbai - 400077

Abstract

In todays world, forecasting a movies opening success is a difficult problem, since it does not always depend on its quality only.External factors such as competing movies, time of the year, region impact the box-office sales for the movie opening.In this study, we use linear regression modelling to generate a model for forecasting box office performance of movies using data collected from multiple sites such as wikipedia, boxofficeindia, koimoi etc.The result is based on various factors for assessment of the movie. The proposed input attributes will be actor,actress,director,cast etc. It is used to forecast results such as whether movie will be hit,flop or neutral,movie rank,box office revenue, regions where the movie will do good busines, suggestion for the movies release date to increase profit .

Keywords: *Movie Trailer, hit class, movies box office performance ,flop class, neutral class*

Contents

List of Figures

Chapter 1

Introduction

1.1 Problem Definition

Our project will generate box-office forecasts for a new movie after the movie has been produced, but before it has been released. Forecasting a movies opening success is a difficult problem, since it does not always depend on its quality only. External factors such as competing movies, time of the year, region impact the Box- Office sales for the movie opening. Nevertheless, forecasting a movies opening success in terms of box-office ticket sales is essential for a movie studio, in order to plan its cost and make the work profitable. As a result,We introduce a simple solution for forecasting movie success in terms of financial success.

1.2 Motivation of the thesis

In todays world, forecasting a movies opening success is a difficult problem, since it does not always depend on its quality only.External factors such as competing movies, time of the year, region impact the box-office sales for the movie opening.The average number of movies produced per year is greater than 1000. So to make the movie profitable, it becomes a matter of concern that the movie succeeds. The objective of the proposed system is to develop a model based on various factors for assessment of the movie which will be able to forecast result using linear regression technique. The data will be collected from different sites and preprocessed.

The users are allowed to enter the data and output will be displayed.

1.3 Scope of the thesis

The scope of the system is to design, develop and implement an interactive system that enables its users to get timely result, to develop a website that is accessible to the general public without any login or registration,to develop a database to store, retrieve and display movie ratings, its cast and script etc. It is used to forecast results such as whether movie will be hit,flop or neutral,movie rank,box office revenue, regions where the movie will do good busines, suggestion for the movies release date to increase profit .

1.3.1 Functional Requirements

- Collection of data : Collecting data from various sites such as wikipedia,boxoffice, koimoi etc.
- Preprocessing of data : Normalizing the data (i.e. converting unstructured data to structured data or use unstructured data).
- GUI: User Interface to display the forecasting results will be created.
- Creation of model: Developing a model with the available data which will be able to forecast result using linear regression technique.
- Display the result : Input will be taken from user using our web application and the output will be displayed.

1.3.2 Non-Functional Requirements

- Accessibility: The tool should be designed so that it is accessible by a large audience; hence we are using Microsoft Visual Studio 2008 and MS-SQL Server 2012; this is one of the most common platform for almost all of the professional application development IDE and Database.
- Audit and control: The tool should have login feature for auditing control.

- Backup: Regular backup should be taken at regular interval on Backup servers.
- Configuration management: Proper configuration is required for the tool to run on Internet Information Services

1.4 Salient contribution

With this project, we try to develop a model with the available data which will be able to forecast result using linear regression technique.

1.5 Organization of the Thesis

1. Chapter 2 describes the Literature Review and Other Description.
2. Chapter 3 describes the milestones, deliverables, process model and roles, responsibility of team members, list of tasks to be performed and the resources needed for their completion.
3. Chapter 4 describes the interface requirements, software product features, software system attributes and database requirements.
4. Chapter 5 will provide framework by describing the high level components and architecture, sub-systems,interfaces, database design and algorithm design.
5. Chapter 6 Provides the information regarding implementation of the project
6. Chapter 7 discusses the basic tests approach undertaken along with the test Plan including features to be tested, testing tools and environment and test specifications.
7. Chapter 8 Is the discussion regarding results and efficiency of the project
8. Chapter 9 provides the conclusion and scope for the future work of the project.

Chapter 2

Literature Survey

This chapter specifies all the knowledge about the platform used and also to understand different concepts and algorithms required for the implementation of the project.

2.1 Paper I

2.1.1 Title of the Paper

Prediction of Movies Box Office Performance Using Social Media [1]

2.1.2 Authors

R. Apala, Merin Jose, Supreme Motnam, C.-C. Chan, Kathy J. Liszka, and Federico de Gregorio1

2.1.3 Publication Year - 2013

2.1.4 Published By IEEE

2.1.5 Abstract

In this study, we apply data mining tools to generate interesting patterns for predicting box office performance of movies using data collected from multiple social media and web sources including Twitter, YouTube and the IMDb movie database. The prediction is based on decision factors derived from a historical movie database, followers count from Twitter, and sentiment analysis of YouTube viewers comments. We label the prediction in three classes, Hit, Neutral and Flop using Wekas K-Means clustering tool. Interesting patterns for prediction are generated by Wekas J48. Since our prediction is for movies yet to be released in summer 2013, the performance of the final results will be validated by a follow-up study.

2.2 Paper II

2.2.1 Title of the Paper

the Near-Weekend Ticket Sales Using Web-based External Factors and Box-office Data [2]

2.2.2 Authors

Seonghoon Moon, Suman Bae, Songkuk Kim

2.2.3 Publication Year 2014

2.2.4 Published By IEEE

2.2.5 Abstract

Posting online reviews and rating their satisfaction on purchased products has become an increasingly popular way to share the information for anonymous can-

didates who has interest in purchasing the product. In addition, people leave their interests and near-future purchasing plan on the web such as search history and search query volume. From this phenomenon, the prediction of sales performance is possible in many products by mining the data sets which are left on the web by consumers online activities. In this paper, we focused on the movie ticket sales which word-of-mouth effect is prominent, and our goal is to forecast the sales performance of the near-weekend using box-office data and external factors such as online reviews, star ratings and search volume. For this work, we gather 1.7 million online reviews and movie ratings, and we also gather the daily search volume of movies title for past three years. Using machine learning techniques and linear modeling, we develop a model for high-accuracy predicting of ticket sales on near-future. We also analyze a relationship between ticket sales performance on weekends and box-office data, online reviews, star ratings, and search volume. Through this work, we support to decide the ideal number of screens for a given weekend, thus it contributes to a substantial increase in the rate of profit on movie markets.

2.3 Ideas adopted from the references

2.3.1 Prediction of Movies Box Office Performance Using Social Media.

1. From this paper we analysed the various factors that could be considered for calculating the success rate. The factors are actor, actress and directors popularity ,their twitters follower count, etc. We also Analysed the concept of sentiment analyses from the same Box-Office opening prediction of Movies through Data Mining

2. Data mining algorithms such as min- max normalization, k-means clustering, Naive Bayesian classification algorithms are used.

2.3.2 Predicting the Near-Weekend Ticket Sales Using Web-based External Factors and Box-office Data.

(a) This paper focuses on movie ticket sales to forecast the near-weekend sales using box-office data and other external factors such as start ratings, movie ratings etc.

(b) Linear modeling is used to develop a model which is highly accurate to predict the ticket sales of near future.

Chapter 3

Software Project Management Plan

This document outlines a brief plan about how the project is to be shaped and includes the milestones and deliverables. This document will serve as a guide for developing the product as part of the project. The document contains brief description of the process model we are going to use, the roles and responsibilities of the team members, list of tasks to be performed and the resources needed for their completion. Updates of this document will serve to record the progress of the project.

3.1 INTRODUCTION

3.1.1 Project Overview

The recent literature pays significant attention to the movie industry because of the industrys unique characteristics. The movie industry is a business with a high profile, and a highly variable revenue stream. A single movie can be the difference between millions of dollars of profits or losses for a studio in a given year. Forecasting a movies opening success is a difficult problem,

since it does not always depend on its quality only. External factors such as competing movies, time of the year and even weather influence the success as these factors impact the Box-Office sales for the movie opening. We introduce a simple solution for forecasting movie success in terms of financial success, movie ratings, etc. As a result, this approach will achieve decent estimations, allowing theatre planning to a certain extent, even for small studios.

3.1.2 Project Deliverables

Deliverable	**Structure**	**Standards**	**Approval Needed By**	**Resources Required**
Project Proposal	Document	As defined in project methodology	Project Guide	Microsoft Word
Software Project Management Plan	Document	As defined in project methodology	Project Guide	Latex
Software Requirements Specification	Document	As defined in project methodology	Project Guide	Latex
Software Design Document	Document	As defined in project methodology	Project Guide	Latex
Database Design and Data Collection	Database Files	As defined in project methodology	Project Guide	Data is collected from websites Wikipedia, Koimoi, BoxOfficeIndia etc SQL server for database design, SQL queries.
User Interface	Prototype	As defined in project methodology	Project Guide	Visual Studio.
Testing Document	Document	As defined in project methodology	Project Guide	Latex
Report and Presentation	Document	As defined in project methodology	Project Guide	Latex, Microsoft Power Point.

3.2 PROJECT ORGANIZATION

3.2.1 Software Process Model

We are going to use the Incremental Model for our project In this model the whole requirement is divided into various builds. Multiple development cycles take place here, making the life cycle a multi-waterfall cycle. As our project is divided in modules such as data collection, pre-processing data, creating GUI, testing, and each module will require the previous module to be completed, we are selecting this model. This model allows us to test and review it at the end of each phase to determine if the project is on right path, helping us build the project correctly and reduce the risk of changing requirements. Moreover there will be no phase overlapping in this module thereby allowing us to build a system with overall functionality.

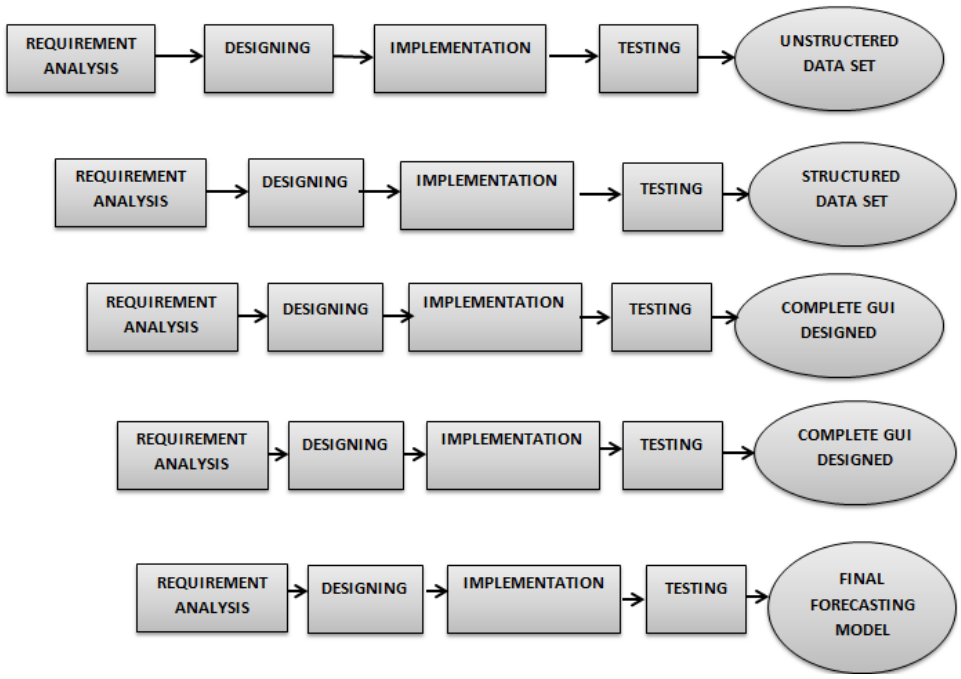


Figure 3.1: Incremental Model

3.2.2 Roles and Responsibilities

The entire project team is responsible for the successful delivery of the product. Each team member will be involved in all the project activities mentioned below:

- Project Plan
- Requirements Specification.
- Documentation and Analysis
- Architecture Specification
- Component/Object Specification
- Designing Database
- Source Code
- Designing User Interface
- Test Plan
- Final Deliverable

3.2.3 Tools and Techniques

3.2.3.1 Software Requirements:

- For front end we use Visual Studio.
- Microsoft Power Point for presentations.
- Microsoft Word and Latex for documentation.
- For back end wel use MS SQL SERVER.
- Macro enabled Microsoft Excel for pre-processing data

3.2.3.2 Hardware Requirements:

- Processor i3 and above.
- Minimum 2GB RAM
- Hard Disk 250GB and above

3.3 PROJECT MANAGEMENT PLAN

3.3.1 Tasks

(a) Software Requirements Specification

(b) Software Design Document

(c) Module 1 - Collection of data.

(d) Module 2- Preprocessing of data.

(e) Module 3- Designing GUI

(f) Module 4- Creation of model and coding to forecast results specified as follows:

- Movies verdict in terms of: Flop, Semi-hit, Super-hit and Block buster
- Its box office revenue prediction.
- Regions where the movie will do good business.
- Suggestion for the movies release month.
- Rating of the movie.

(g) Module 5-Testing

3.3.1.1 SRS(Software Requirement Specification)

3.3.1.1.1 Description

This document will completely describe all the functions of a proposed system and the constraints under which it must operate.

3.3.1.1.2 Deliverables and Milestones

An SRS document which will be reviewed by the project Guide and finalised, the revised SRS will be a milestone.

3.3.1.1.3 Resources Needed

Latex for documentation and Internet will be the basic requirements.

3.3.1.1.4 Risks and Contingencies

Change in requirements will lead to change in SRS.

3.3.1.2 Software Design Document

3.3.1.2.1 Description

This document will consist a detailed written description of software product. It will demonstrate how the design will accomplish the functional and non-functional requirements specified in the SRS .It will give us a view of system architecture, database design, interface of the project, and algorithm design.

3.3.1.2.2 Deliverables and Milestones

The document will be reviewed by the project Guide and finalized, the revised document will be a milestone.

3.3.1.2.3 Resources Needed

Microsoft Word, Latex for documentation and Internet will be the basic requirements.

3.3.1.2.4 Risks and Contingencies

Designing phase cannot start until all the requirements achieved. Change in requirement will result in change in software design.

3.3.1.3 Module 1-Collection of data

3.3.1.3.1 Description

Data is collected from various websites like Wikipedia, Koimoi, Boxoffice-india and Bollywood hungama. We designed a structure for our database. Some of the parameters collected for our database are as follows:

- Actor and actresses rank in the industry.

- Details of various movies, which include the movie name, genre of the movie, cast, director and the release date of the movie.
- Box-office revenue of the movie.
- Month of Release.
- Region-wise revenue earned by movies.

3.3.1.3.2 Deliverables and Milestones

The deliverable will be database with unstructured data stored in it. The task of populating the database will be carried regularly.

3.3.1.3.3 Resources Needed

We will require SQL server for storing the database, which will be integrated with UI using Visual studio.

3.3.1.3.4 Dependencies and Constraints

Changes in the results to be implemented will result in changing the dataset and factors to be considered in dataset.

3.3.1.4 Module 2- Pre-processing of data.

3.3.1.4.1 Description

Macro enabled workbooks are created using VBA coding to pre-process the data. Here the data collected from various social media sites will be converted to structured format using the workbooks.

3.3.1.4.2 Deliverables and Milestones

The deliverable will be a structured dataset which can then be used for implementation of next module.

3.3.1.4.3 Resources Needed

Microsoft Excel to create workbooks. SQL server for storing the database, which will be integrated with UI using Visual studio.

3.3.1.4.4 Dependencies and Constraints

Changes in the results to be implemented will result in changing the dataset.

3.3.1.5 Module 3- Designing GUI

3.3.1.5.1 Description

An interface will be developed in which the user can choose the various options for displaying the output results. Admin will have login id and password. He/she can insert data in database using the interface. The GUI should be user compatible.

3.3.1.5.2 Deliverables and Milestones

The User Interface after integration of all the features will become a milestone.

3.3.1.5.3 Resources Needed

Visual Studio, .Net framework, C sharp coding

3.3.1.5.4 Dependencies and Constraints

This task does not depend on any module.

3.3.1.6 Module 4- Creation of Model.

3.3.1.6.1 Description

Regression based modelling technique will be used as an algorithm for creating a model which will forecast the required results.

3.3.1.6.2 Deliverables and Milestones

Source code of the algorithm been used.

3.3.1.6.3 Resources Needed

Visual Studio, SQL server.

3.3.1.6.4 Dependencies and Constraints

Changes in dataset may affect the code.

3.3.1.7 Module 5- Testing.

3.3.1.7.1 Description

It deals with testing various modules and applications of the project using various testing techniques like black box testing, white box testing, etc.

3.3.1.7.2 Deliverables and Milestones

The deliverable will be a list of features not working with the changes been incorporated. The final interface after incorporating the changes and resolving errors will be a milestone.

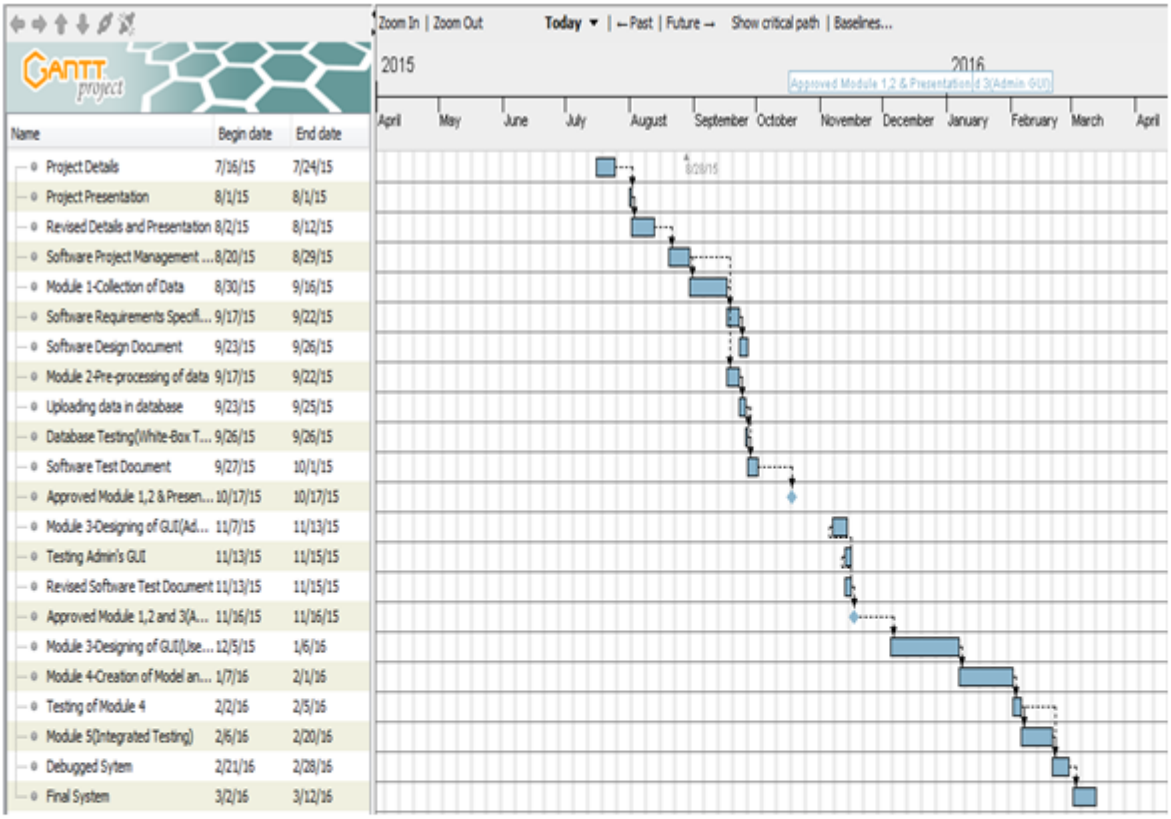
3.3.1.7.3 Resources Needed

Test Cases

3.3.1.7.4 Dependencies and Constraints

This task can be performed only when all the above tasks are completed.

3.3.2 Timetable



Name	Begin date	End date
Project Details	7/16/15	7/24/15
Project Presentation	8/1/15	8/1/15
Revised Details and Presentation	8/2/15	8/12/15
Software Project Management ...	8/20/15	8/29/15
Module 1-Collection of Data	8/30/15	9/16/15
Software Requirements Specifi...	9/17/15	9/22/15
Software Design Document	9/23/15	9/26/15
Module 2-Pre-processing of data	9/17/15	9/22/15
Uploading data in database	9/23/15	9/25/15
Database Testing(White-Box T...	9/26/15	9/26/15
Software Test Document	9/27/15	10/1/15
Approved Module 1,2 & Presen...	10/17/15	10/17/15
Module 3-Designing of GUI(Adi...	11/7/15	11/13/15
Testing Admin's GUI	11/13/15	11/15/15
Revised Software Test Document	11/13/15	11/15/15
Approved Module 1,2 and 3(A...	11/16/15	11/16/15
Module 3-Designing of GUI(Use...	12/5/15	1/6/16
Module 4-Creation of Model an...	1/7/16	2/1/16
Testing of Module 4	2/2/16	2/5/16
Module 5(Integrated Testing)	2/6/16	2/20/16
Debugged Sytem	2/21/16	2/28/16
Final System	3/2/16	3/12/16

Figure 3.2: Gantt Chart

Task	**Start Date**	**End Date**
Project Details	16/7/15	24/7/15
Project Presentation	1/8/15	1/8/15
Revised Details and Presentation	2/8/15	12/8/15
Software Project Management Plan	20/8/15	29/8/15
Module 1-Collection of Data	30/8/15	16/9/15
Software Requirements Specification	17/9/15	22/9/15
Software Design Document	23/9/15	26/9/15
Module 2-Pre-processing of data	17/9/15	22/9/15
Uploading data in database	23/9/15	25/9/15
Database Testing(White-box Testing)	26/9/15	26/9/15
Software Test Document	27/9/15	1/10/15
Approved Module 1 and 2	17/10/15	17/10/15
Module 3- Designing GUI(Admin Side)	7/11/15	13/11/15
Testing Admin GUI	13/11/15	15/11/15
Revised Software Test Document	13/11/15	15/11/15
Approved Module 1,2 and 3(Admin Side)	16/11/15	16/11/15
Module 3-Designing GUI(User's Side)	5/12/15	6/1/16
Module 4-Creation of model and coding to forecast results	7/1/16	1/2/16
Testing of Module 4	2/2/16	5/2/16
Module 5-Integrated Testing	5/2/16	20/2/16
Debugged system	21/2/16	28/2/16
Final system	2/3/16	12/3/16

Chapter 4

Software Requirements Specification

A software requirement specification defines how an application will interact with system hardware, other programs and human users in a wide variety of real-world situations. Parameters such as operating speed, response time, availability, portability, maintainability, footprint, security and speed of recovery from adverse events are evaluated. This chapter discusses about all of the functions of a proposed system and the constraints under which it must operate. The document contains brief description of the product, external interface requirements, software product features, software system attributes and database requirements.

4.1 Introduction

4.1.1 Product Overview

MovieBuzz is aimed towards providing a simple solution for forecasting movie success in terms of financial success and viewer opinion. The product will be a system proposing outputs such as will the movie be flop, semi-hit, super-hit or block-buster, movie rank, box office revenue prediction ,region where the movie will do good business, suggestion for the movies release month. with regression based modeling. It collects data from sites such as wikipedia [3],

boxofficeindia.com , koimoi.com [4, 5] etc.The system should be user-friendly and reliable for the above purpose.

4.2 Specific Requirements

4.2.1 External Interface Requirements

The external interface requirements contains user interfaces and communication protocols.

4.2.1.1 User Interfaces

4.2.1.1.1 Admin login page containing upload module

The upload module will allow the system administrator to upload new data to the existing database. It will be useful for the administrator to upload data collected from various sites that will be preprocessed and uploaded with current data.The system administrator will be given a separate login page where he/she can login and can upload the data from excel sheet to the database. A normal user can be able to explore the website without any login or registration.

4.2.1.1.2 User web page to forecast below mentioned outputs

- Will the movie be flop, semi-hit, super-hit or block-buster by considering various attributes like movie name,actor name,actress name,director name and genre collected from various websites.
- Movies box office revenue prediction..
- Region where the movie will do good business by considering its revenue generation in different parts of the country.
- Suggestion for the movies release month
- Rating of the movie. .

4.2.1.1.3 Output Screen

The webpage will contain textboxes for inputs like movie name,director name,actor name,actress name and drop-down list for genre and release month.After clicking on submit, all the above mentioned outputs will be displayed to user.The output window will display all the inputs entered by user along with parameter contribution of each input. The user can reset the form, if required, by clicking on the reset form button.

4.2.1.2 Communications Protocols

Our project will require HTTP communication protocol as it is web related. It is an request-response client-server protocol. An http client sends a request message to an http server. The server, in turn, returns a response message.

4.2.2 Software Product Features

4.2.2.1 Administrator login

A normal user can be able to explore the website without any login or registration. The system administrator will be given a separate login page where he/she can login and can upload the data from excel sheet to the database.

4.2.2.2 Web page for admin to upload data to database

This component will allow the system administrator to upload new data to the existing database.This component will be useful for the administrator to upload data collected from various sites that will be preprocessed and uploaded with current data.

4.2.2.3 Entry of inputs and selection of outputs from user

The users are requested to enter the details like movie name,actor name,actress name,director name in textboxes. He/She will be able to select the genre

and release month from drop-down list. After clicking on submit button ,the user will be able to see the forecasting results as specified follows:

4.2.2.3.1 Flop, Semi-hit, Super-hit and Block-buster forecast

This feature will consider various attributes and information pertaining to movie details,genre,release month,actor,actress and directors details collected from various websites, that will be useful to determine movie's performance that whether it will be flop, semi-hit, super-hit or block-buster.

4.2.2.3.2 Movie box office revenue prediction

This feature will help in deciding the estimated revenue that will be generated by the movie at box-office. This will provide estimate in terms of percentage revenue contribution of the movie to boxoffice.

4.2.2.3.3 Region where the movie will do good business

This feature will have information regarding the movies region wise performance by considering its revenue generation in different parts of the country.This will help in predicting the region where the movie will do good business.

4.2.2.3.4 Movie's release month

This feature will provide the suggestion for the movies release month by considering attributes such as movie details,genre,release month, actor,actress and directors details collected from various websites.

4.2.2.3.5 Movie Rating

This feature will be useful to determine movie's rating by considering various attributes such as movie details,genre,release month, actor,actress and directors details collected from various websites.

4.2.2.4 Display the result

After clicking on submit button,the result will be displayed to user in the output window.The output window will display all the inputs entered by user along with parameter contribution of each input.

4.2.3 Software System Attributes

4.2.3.1 Availability

The database should be backed up regularly to prevent loss of data.The system should be available to the user 24x7.

4.2.3.2 Maintainability

Filmfare award winning actors, actresses and directors were added to the database. New movie name can be be added to the system.

4.2.3.3 Performance

The system should be checked for bugs and errors at regular intervals in order to increase accuracy and performance.

4.2.4 Database Requirements

The database will comprises of information about all the aforementioned attributes needed to be created in order to carry out further modelling or processing. Excel will be used to handle the large number of raw and unstructured data records. That data will be processed and structured and then fed to the SQL database tool.The information available on the excel will be aggregated into a relational database and searching for information will be done using SQL queries. The system will be integrated with database using .NET framework data provider for SQL server.

Chapter 5

Software Design Description

The software design document would demonstrate how the design will accomplish the functional and non-functional requirements captured in the SRS. The document will provide framework by describing the high level components and architecture, sub-systems,interfaces, database design and algorithm design.This is achieved through the use of architectural patterns, design patterns and user interfaces.

5.1 Design Overview

MovieBuzz system will be developed using mainly client-server architecture. The feature of this architecture allows configurality,authentication; inclusion of programs that handle and generate dynamic content; module support etc. Which will allow development of the application with added functionalities over time.

5.1.1 Requirements Traceability Matrix

	Admin component	Movie details , Actor,Actress and Director details	Box Office Net Gross	Region wise box office data
Upload Module	X			
Blockbuster, Super-Hit, Semi-Hit and Flop Forecast		X	X	
Movie Rating		X	X	
Projected Box Office Revenue			X	
Region where the movie will do good business		X		X
Movie Release Month		X	X	
Suggestion		X	X	

5.2 System Architectural Design:

The chosen architecture for our project is Client/Server. The server houses and provides high-end, computing-intensive services to the client on demand. These services can include applications access, storage, file sharing, printer access and/or direct access to the servers raw computing power. Client/server architecture works when the client computer sends a resource or process request to the server over the network connection, which is then processed and delivered to the client. A server computer can manage several clients simultaneously, whereas one client can be connected to several servers at a time, each providing a different set of services. In its simplest form, the Internet is also based on client/server architecture where the Web server serves many simultaneous users with Web page and or website data.The risk associated this architecture is network failure.

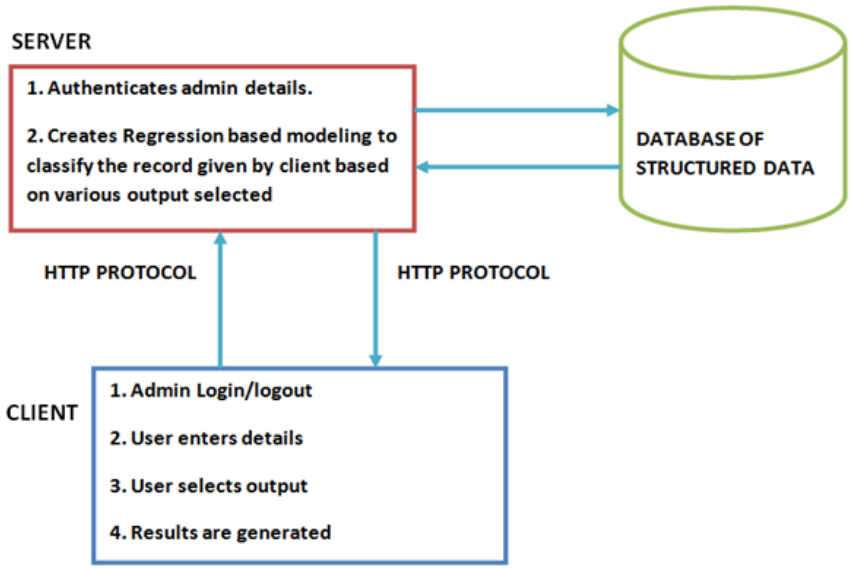


Figure 5.1: client-server architecture

5.2.1 Alternate Designs

Alternate designs that can be used is REST architecture.Due to complexity of the REST architecture we have preferred Client-Server architecture over REST.

5.2.2 System Interface Design

The files that will be used with the project will be excel sheets containing raw and unstructured data. That will be processed and structured and then fed to the database.The system will be integrated with database using .NET Framework Data provider for SQL Server.

5.3 DETAILED COMPONENT DESCRIPTION

5.3.1 Admin Component

This component will allow the system administrator to upload new data to the existing database.It will be useful for the administrator to upload data collected from various sites that will be preprocessed and uploaded.

5.3.2 Movie details, Actor,Actress and Director details

This component has various attributes and information pertaining to Movie details,Actor,Actress and Director's details collected from various websites that will be useful to determine movie's ranking,it's performance whether it will be hit,flop or neutral.

5.3.3 Box Office Gross

This component will help in deciding the estimated revenue that will be generated by the movie at box-office.It has the estimates regarding the gross income earned by number of movies that will help in creation of model that will try to provide accurate results.

5.3.4 Region wise box office data

This component will have information regarding the movie's region wise performance by considering it's revenue generation in different parts of the country.This will help in predicting the region where the movie will do good business.

5.4 USER INTERFACE DESIGN

5.4.1 UI layout for user and Admin

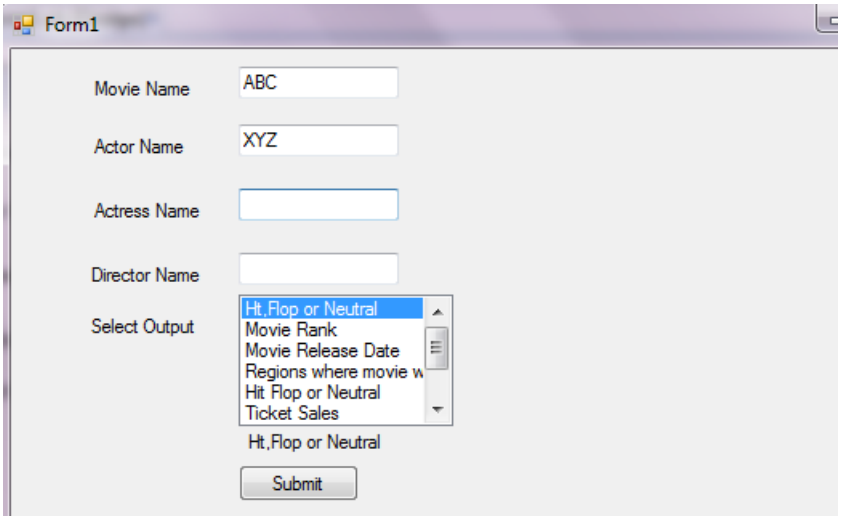


Figure 5.2: proposed layout for user

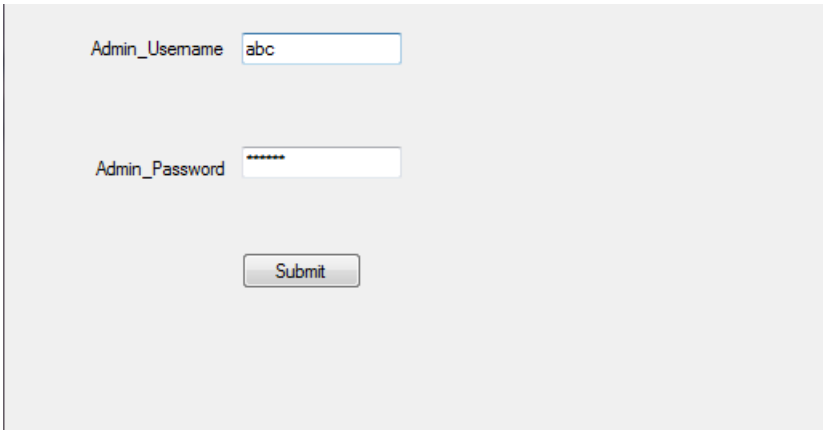


Figure 5.3: separate login page for admin

5.4.2 Class Diagram, Data Flow Diagram,Activity Diagram,UseCase Diagram

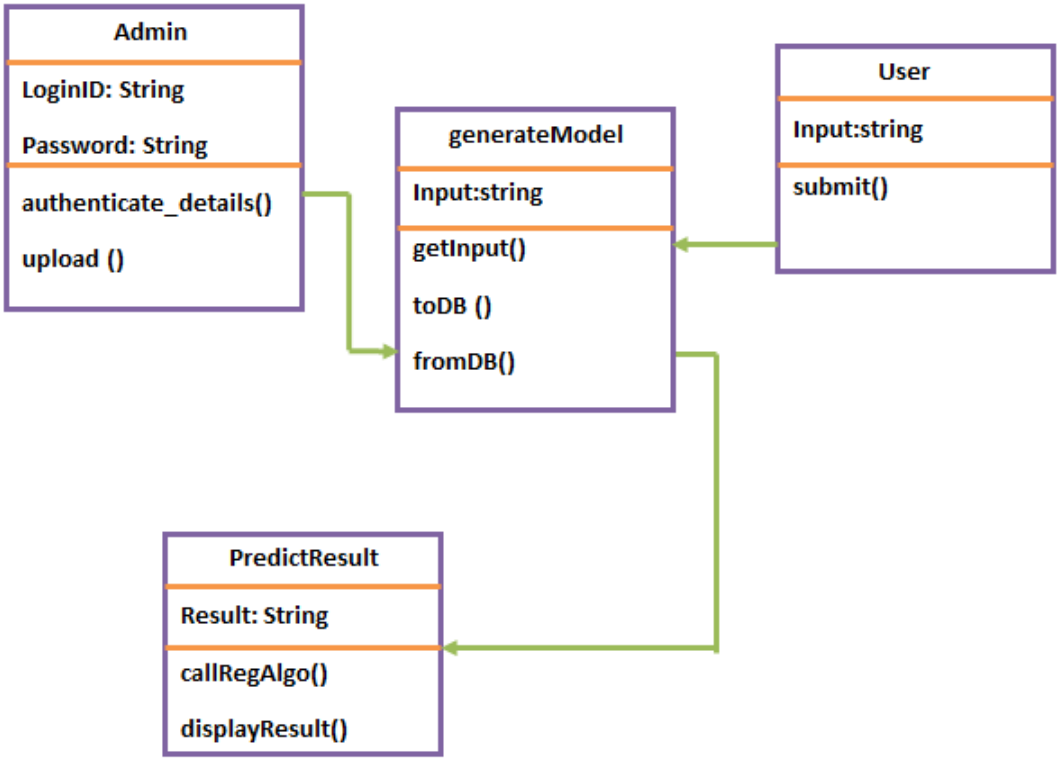


Figure 5.4: interaction between different objects

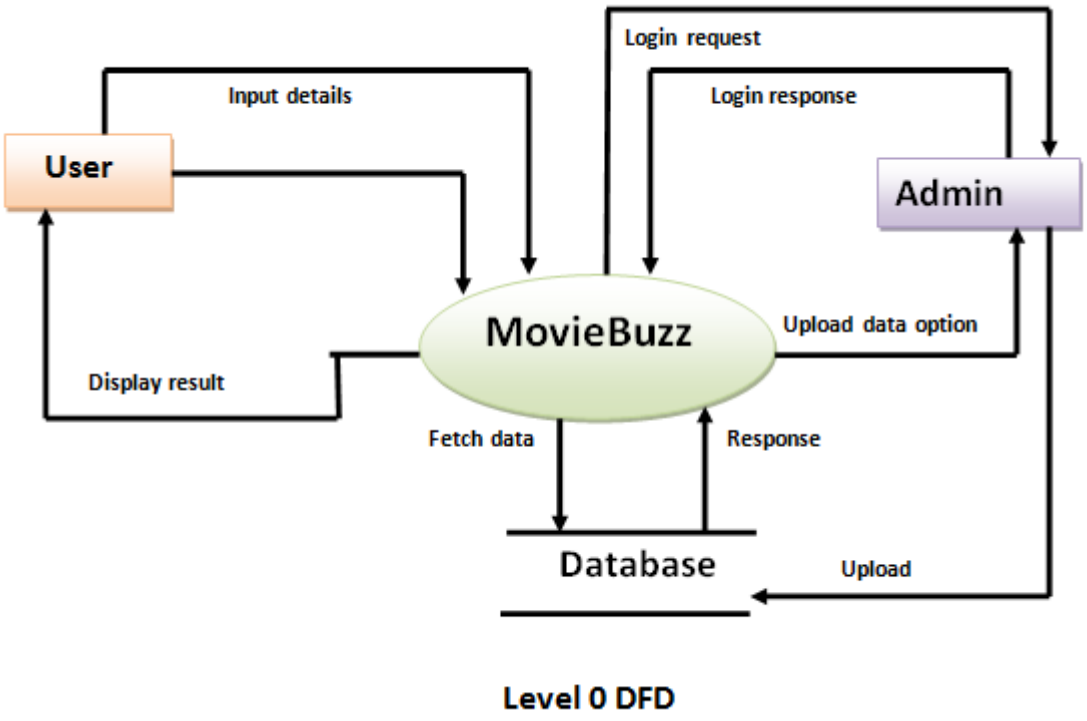


Figure 5.5: data flow in the system

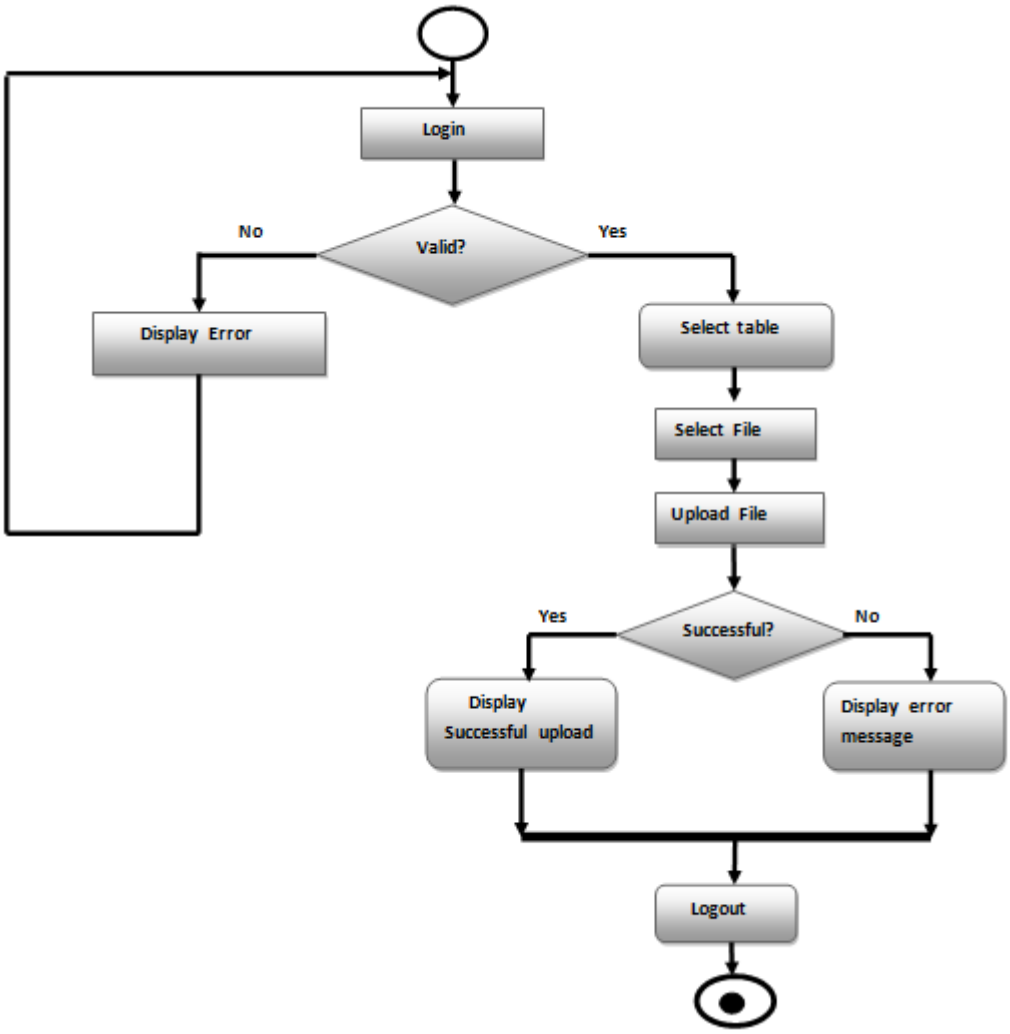


Figure 5.6: proposed activity diagram

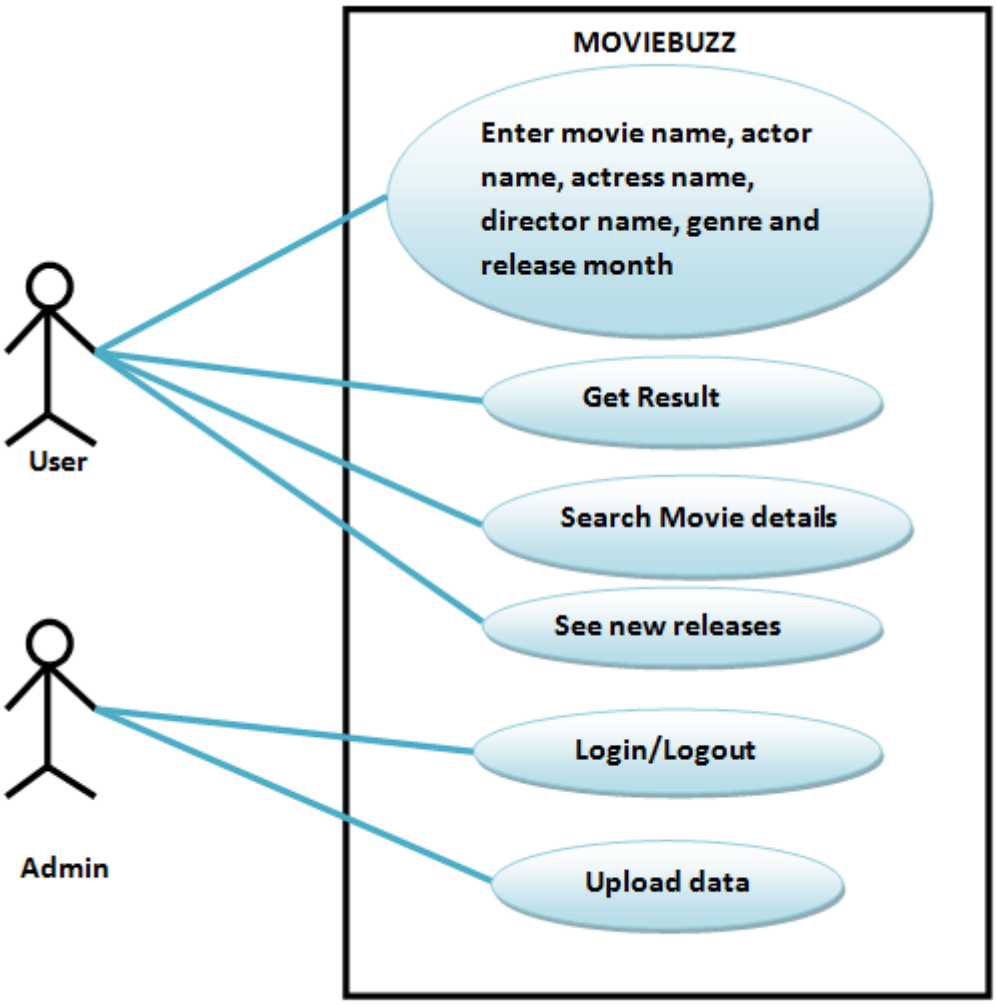


Figure 5.7: various functionalities given to user and admin

Chapter 6

Implementation

This chapter discusses about the implementation of the MovieBuzz system.

6.1 Technologies Used

(a) Front end - Visual Studio C # .NET with ASP .NET.

(b) Back end - MS SQL SERVER 2012.

(c) Pre processing -Macro Enabled Microsoft excel.

(d) Documentation -Latex and Microsoft word

6.2 Detailed Implementation

6.2.1 Data Collection

(a) Data required for our project is collected from websites such as Wikipedia,box office.com, Bollywood hungama [6] and koimoi.com.

(b) Some links from where the data is collected are as follows:

i. https://en.wikipedia.org/wiki/List_of_Bollywood_films_of_2015

ii. http://www.koimoi.com/box-office-bollywood-films-of-2013/

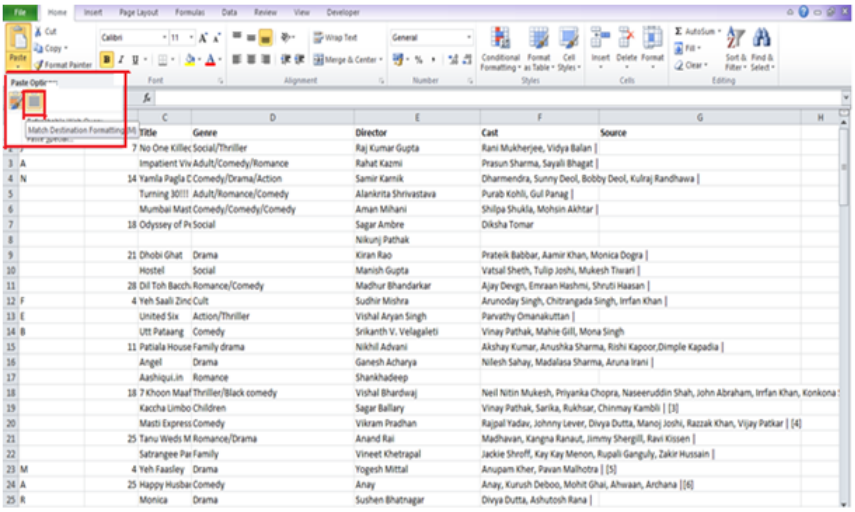
iii. http://boxofficeindia.com/Collections/net$_b ox_o f fice/delhi_u p/\#.Vg5fpS6qqkp$

(c) The example of data available on https://en.wikipedia.org/wiki/List_-of_Bollywood_films_of_2013:

Opening		Title	Genre	Director	Cast	Source
JANUARY	4	*Table No. 21*	Thriller/Drama	Aditya Datt	Paresh Rawal, Rajeev Khandelwal, Tena Desae	[4]
		Dehraadun Diary	Thriller	Milind Ukey	Adhyayan Suman, Rohit Bakshi (actor), Ragini Nandwani, Ashwini Kalsekar, Rati Agnihotri	[5]
		Meri Shadi Karao	Comedy	Syed Noor	Gurdeep Singh Mehndi	[6]
	11	*Matru Ki Bijlee Ka Mandola*	Comedy	Vishal Bhardwaj	Imran Khan, Anushka Sharma, Pankaj Kapur	[7]
		Pyaar Mein Aisa Hota Hai	Thriller	Rajeesh Kumar	Mohd Nazim, Aanamika Bawa, Rohit Rajawat, Aayushi Sharma	[8]
		Gangoobai	Drama	Priya Krishnaswamy	Sarita Joshi, Purab Kohli, Mita Vasisht	[9]
	18	*Inkaar*	Romance/Crime	Sudhir Mishra	Arjun Rampal, Chitrangada Singh, Gaurav Dwivedi	[10]
		Mumbai Mirror	Action	Ankush Bhatt	Sachiin Joshi, Gihana Khan, Prakash Raj	[11]
		Bandook	Crime	Aditya Om	Aditya Om, Manisha Kelkar, Arshad Khan	[12]
	25	*Race 2*	Action/Thriller	Abbas-Mustan	Saif Ali Khan, Deepika Padukone, John Abraham, Jacqueline Fernandez, Anil Kapoor, Ameesha Patel	[13]
		Akaash Vani	Romance	Luv Ranjan	Kartik Tiwari, Nushrat Bharucha	[14]
		Main Krishna Hoon	Animation/Musical	Rajiv S. Ruia	Hrithik Roshan, Juhi Chawla, Rajneesh Duggal, Misti Mukherjee, Katrina Kaif	[15]
F	1	*David*	Crime	Bejoy Nambiar	Vikram, Lara Dutta, Vinay Virmani, Neil Nitin Mukesh, Tabu	[16]
		Vishwaroop	Action	Kamal Hassan	Kamal Hassan, Pooja Kumar, Shekhar Kapur, Andrea Jeremiah, Rahul Bose	[17]
		Mai...	Drama/Family	Mahesh Kodiyal	Asha Bhosle, Ram Kapoor, Padmini Kolhapure	[18]
				Avinash		[19]

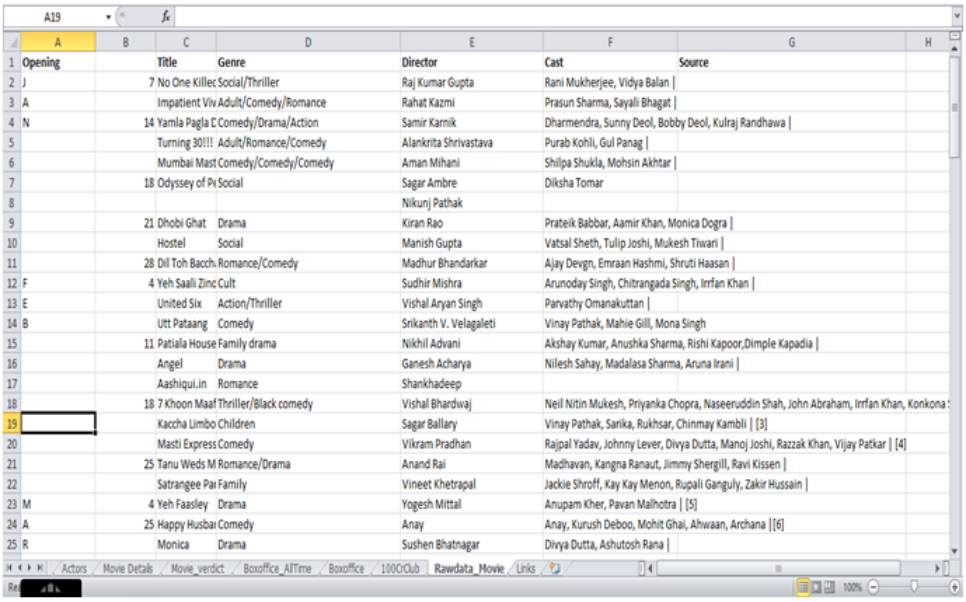
Figure 6.1: data available on https://en.wikipedia.org/wiki/List_of_Bollywood_-films_of_2013

(d) After selecting, the data it is pasted in excel sheet as follows:



	C	D	E	F	G	H
	Title	Genre	Director	Cast	Source	
	7 No One Kille	Social/Thriller	Raj Kumar Gupta	Rani Mukherjee, Vidya Balan |		
A	Impatient Viv	Adult/Comedy/Romance	Rahat Kazmi	Prasun Sharma, Sayali Bhagat |		
N	14 Yamla Pagla D	Comedy/Drama/Action	Samir Karnik	Dharmendra, Sunny Deol, Bobby Deol, Kulraj Randhawa |		
	Turning 30!!!	Adult/Romance/Comedy	Alankrita Shrivastava	Purab Kohli, Gul Panag |		
	Mumbai Mast	Comedy/Comedy/Comedy	Aman Mihani	Shilpa Shukla, Mohsin Akhtar |		
	18 Odyssey of Pe	Social	Sagar Ambre	Diksha Tomar		
			Nikunj Pathak			
	21 Dhobi Ghat	Drama	Kiran Rao	Prateik Babbar, Aamir Khan, Monica Dogra |		
	Hostel	Social	Manish Gupta	Vatsal Sheth, Tulip Joshi, Mukesh Tiwari |		
	28 Dil Toh Bacch	Romance/Comedy	Madhur Bhandarkar	Ajay Devgn, Emraan Hashmi, Shruti Haasan |		
F	4 Yeh Saali Zind	Cult	Sudhir Mishra	Arunoday Singh, Chitrangada Singh, Irrfan Khan |		
E	United Six	Action/Thriller	Vishal Aryan Singh	Parvathy Omanakuttan |		
B	Utt Pataang	Comedy	Srikanth V. Velagaleti	Vinay Pathak, Mahie Gill, Mona Singh		
	11 Patiala House	Family drama	Nikhil Advani	Akshay Kumar, Anushka Sharma, Rishi Kapoor,Dimple Kapadia |		
	Angel	Drama	Ganesh Acharya	Nilesh Sahay, Madalasa Sharma, Aruna Irani |		
	Aashiqui.in	Romance	Shankhadeep			
	18 7 Khoon Maaf	Thriller/Black comedy	Vishal Bhardwaj	Neil Nitin Mukesh, Priyanka Chopra, Naseeruddin Shah, John Abraham, Irrfan Khan, Konkona		
	Kaccha Limbo	Children	Sagar Ballary	Vinay Pathak, Sarika, Rukhsar, Chinmay Kambli | [3]		
	Masti Express	Comedy	Vikram Pradhan	Rajpal Yadav, Johnny Lever, Divya Dutta, Manoj Joshi, Razzak Khan, Vijay Patkar | [4]		
	25 Tanu Weds M	Romance/Drama	Anand Rai	Madhavan, Kangna Ranaut, Jimmy Shergill, Ravi Kissen |		
	Satrangee Par	Family	Vineet Khetrapal	Jackie Shroff, Kay Kay Menon, Rupali Ganguly, Zakir Hussain |		
M	4 Yeh Faasley	Drama	Yogesh Mittal	Anupam Kher, Pavan Malhotra | [5]		
A	25 Happy Husbar	Comedy	Anay	Anay, Kurush Deboo, Mohit Ghai, Ahwaan, Archana |[6]		
R	Monica	Drama	Sushen Bhatnagar	Divya Dutta, Ashutosh Rana |		

Figure 6.2: copy pasted data



	A	B	C	D	E	F	G	H
1	Opening		Title	Genre	Director	Cast	Source	
2	J		7	No One Kille Social/Thriller	Raj Kumar Gupta	Rani Mukherjee, Vidya Balan |		
3	A			Impatient Viv Adult/Comedy/Romance	Rahat Kazmi	Prasun Sharma, Sayali Bhagat |		
4	N		14	Yamla Pagla D Comedy/Drama/Action	Samir Karnik	Dharmendra, Sunny Deol, Bobby Deol, Kulraj Randhawa |		
5				Turning 30!!! Adult/Romance/Comedy	Alankrita Shrivastava	Purab Kohli, Gul Panag |		
6				Mumbai Mast Comedy/Comedy/Comedy	Aman Mihani	Shilpa Shukla, Mohsin Akhtar |		
7			18	Odyssey of Pe Social	Sagar Ambre	Diksha Tomar		
8					Nikunj Pathak			
9			21	Dhobi Ghat Drama	Kiran Rao	Prateik Babbar, Aamir Khan, Monica Dogra |		
10				Hostel Social	Manish Gupta	Vatsal Sheth, Tulip Joshi, Mukesh Tiwari |		
11			28	Dil Toh Bacch Romance/Comedy	Madhur Bhandarkar	Ajay Devgn, Emraan Hashmi, Shruti Haasan |		
12	F		4	Yeh Saali Zind Cult	Sudhir Mishra	Arunoday Singh, Chitrangada Singh, Irrfan Khan |		
13	E			United Six Action/Thriller	Vishal Aryan Singh	Parvathy Omanakuttan |		
14	B			Utt Pataang Comedy	Srikanth V. Velagaleti	Vinay Pathak, Mahie Gill, Mona Singh		
15			11	Patiala House Family drama	Nikhil Advani	Akshay Kumar, Anushka Sharma, Rishi Kapoor,Dimple Kapadia |		
16				Angel Drama	Ganesh Acharya	Nilesh Sahay, Madalasa Sharma, Aruna Irani |		
17				Aashiqui.in Romance	Shankhadeep			
18			18	7 Khoon Maaf Thriller/Black comedy	Vishal Bhardwaj	Neil Nitin Mukesh, Priyanka Chopra, Naseeruddin Shah, John Abraham, Irrfan Khan, Konkona		
19				Kaccha Limbo Children	Sagar Ballary	Vinay Pathak, Sarika, Rukhsar, Chinmay Kambli | [3]		
20				Masti Express Comedy	Vikram Pradhan	Rajpal Yadav, Johnny Lever, Divya Dutta, Manoj Joshi, Razzak Khan, Vijay Patkar | [4]		
21			25	Tanu Weds M Romance/Drama	Anand Rai	Madhavan, Kangna Ranaut, Jimmy Shergill, Ravi Kissen |		
22				Satrangee Par Family	Vineet Khetrapal	Jackie Shroff, Kay Kay Menon, Rupali Ganguly, Zakir Hussain |		
23	M		4	Yeh Faasley Drama	Yogesh Mittal	Anupam Kher, Pavan Malhotra | [5]		
24	A		25	Happy Husbar Comedy	Anay	Anay, Kurush Deboo, Mohit Ghai, Ahwaan, Archana |[6]		
25	R			Monica Drama	Sushen Bhatnagar	Divya Dutta, Ashutosh Rana |		

Figure 6.3: Raw data

6.2.2 Data Pre-processing

Data collected is pre-processed using macro enabled Excel Workbooks.We are doing so, because websites keep changing their format of webpages without informing any of its users, so complete dependency of our data on any website will result in affecting our database.
some options given in macro are as follows:

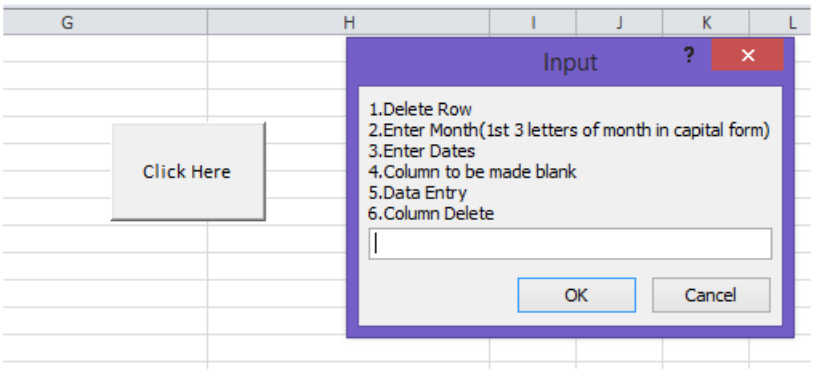
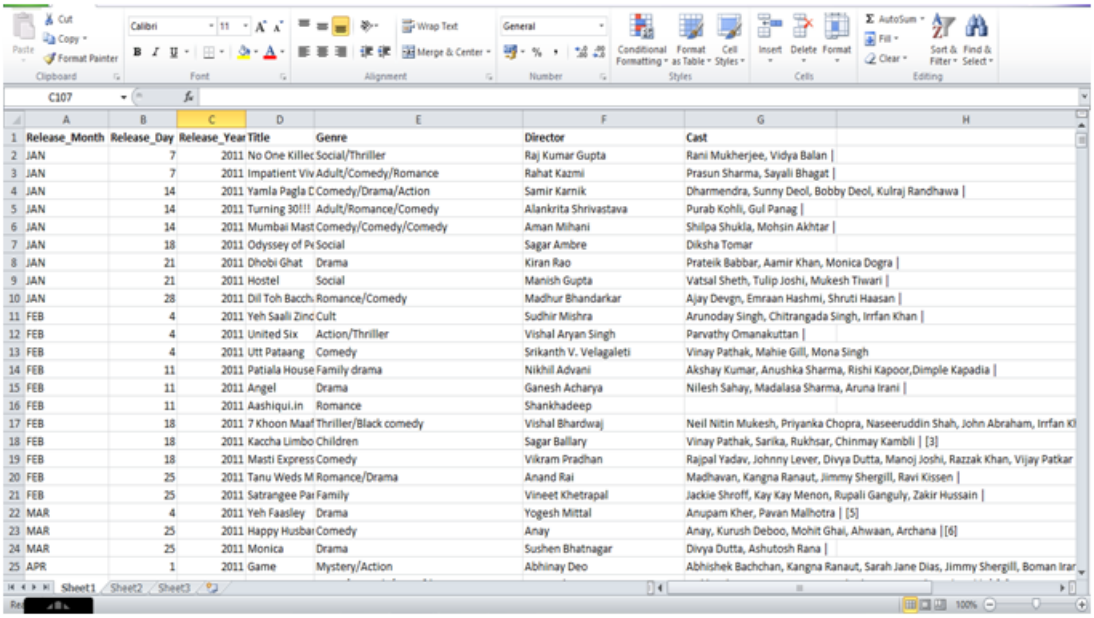


Figure 6.4: Macro options

Pre-processed data is as below



	Release_Month	Release_Day	Release_Year	Title	Genre	Director	Cast
2	JAN	7	2011	No One Killec	Social/Thriller	Raj Kumar Gupta	Rani Mukherjee, Vidya Balan |
3	JAN	7	2011	Impatient Viv	Adult/Comedy/Romance	Rahat Kazmi	Prasun Sharma, Sayali Bhagat |
4	JAN	14	2011	Yamla Pagla D	Comedy/Drama/Action	Samir Karnik	Dharmendra, Sunny Deol, Bobby Deol, Kulraj Randhawa |
5	JAN	14	2011	Turning 30!!!	Adult/Romance/Comedy	Alankrita Shrivastava	Purab Kohli, Gul Panag |
6	JAN	14	2011	Mumbai Mast	Comedy/Comedy/Comedy	Aman Mihani	Shilpa Shukla, Mohsin Akhtar |
7	JAN	18	2011	Odyssey of Pe	Social	Sagar Ambre	Diksha Tomar
8	JAN	21	2011	Dhobi Ghat	Drama	Kiran Rao	Prateik Babbar, Aamir Khan, Monica Dogra |
9	JAN	21	2011	Hostel	Social	Manish Gupta	Vatsal Sheth, Tulip Joshi, Mukesh Tiwari |
10	JAN	28	2011	Dil Toh Bacch	Romance/Comedy	Madhur Bhandarkar	Ajay Devgn, Emraan Hashmi, Shruti Haasan |
11	FEB	4	2011	Yeh Saali Zinc	Cult	Sudhir Mishra	Arunoday Singh, Chitrangada Singh, Irrfan Khan |
12	FEB	4	2011	United Six	Action/Thriller	Vishal Aryan Singh	Parvathy Omanakuttan |
13	FEB	4	2011	Utt Pataang	Comedy	Srikanth V. Velagaleti	Vinay Pathak, Mahie Gill, Mona Singh
14	FEB	11	2011	Patiala House	Family drama	Nikhil Advani	Akshay Kumar, Anushka Sharma, Rishi Kapoor,Dimple Kapadia |
15	FEB	11	2011	Angel	Drama	Ganesh Acharya	Nilesh Sahay, Madalasa Sharma, Aruna Irani |
16	FEB	11	2011	Aashiqui.in	Romance	Shankhadeep	
17	FEB	18	2011	7 Khoon Maaf	Thriller/Black comedy	Vishal Bhardwaj	Neil Nitin Mukesh, Priyanka Chopra, Naseeruddin Shah, John Abraham, Irrfan Kl
18	FEB	18	2011	Kaccha Limbo	Children	Sagar Ballary	Vinay Pathak, Sarika, Rukhsar, Chinmay Kambli | [3]
19	FEB	18	2011	Masti Express	Comedy	Vikram Pradhan	Rajpal Yadav, Johnny Lever, Divya Dutta, Manoj Joshi, Razzak Khan, Vijay Patkar
20	FEB	25	2011	Tanu Weds M	Romance/Drama	Anand Rai	Madhavan, Kangna Ranaut, Jimmy Shergill, Ravi Kissen |
21	FEB	25	2011	Satrangee Par	Family	Vineet Khetrapal	Jackie Shroff, Kay Kay Menon, Rupali Ganguly, Zakir Hussain |
22	MAR	4	2011	Yeh Faasley	Drama	Yogesh Mittal	Anupam Kher, Pavan Malhotra | [5]
23	MAR	25	2011	Happy Husbar	Comedy	Anay	Anay, Kurush Deboo, Mohit Ghai, Ahwaan, Archana |[6]
24	MAR	25	2011	Monica	Drama	Sushen Bhatnagar	Divya Dutta, Ashutosh Rana |
25	APR	1	2011	Game	Mystery/Action	Abhinay Deo	Abhishek Bachchan, Kangna Ranaut, Sarah Jane Dias, Jimmy Shergill, Boman Irar

Figure 6.5: Pre processed data

6.2.2.1 Designing Database

The database is designed and created using SQL server. After creating the database the pre-processed data is uploaded in respective tables.

6.2.2.2 GUI

(a) DESIGNING INTERFACE

Visual studio 2008 is used to create user interface.In visual studio an ASP.NET Web App project is created and User interface designed is as follows:

i. Home Page

Figure 6.6: Home Page

ii. Login Page

Figure 6.7: Login Page

iii. after login following page appears:

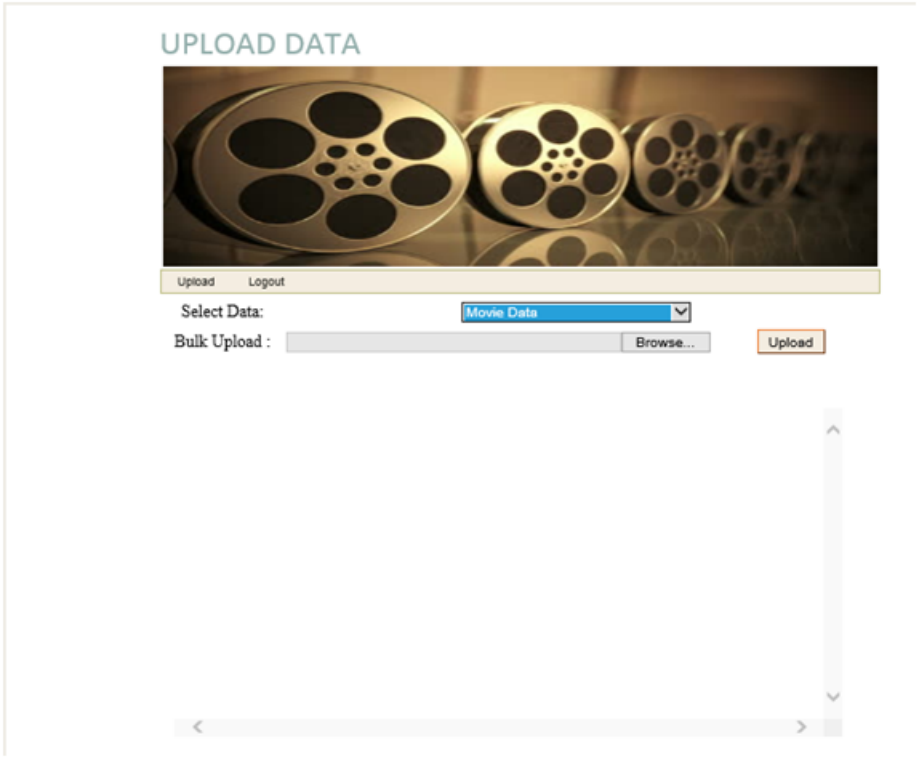


Figure 6.8: After Login page

iv. Uploading data in database Admin can upload data in database from front-end as follows

A. From the drop down menu select the table in which you want to upload the data.

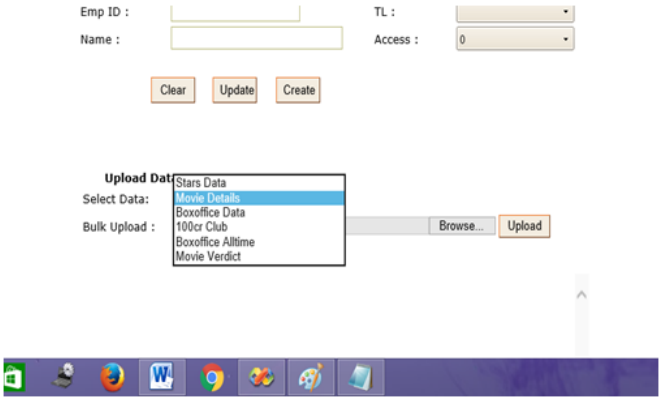


Figure 6.9: Selection of table

B. Click on browse and select the path where your file is stored.

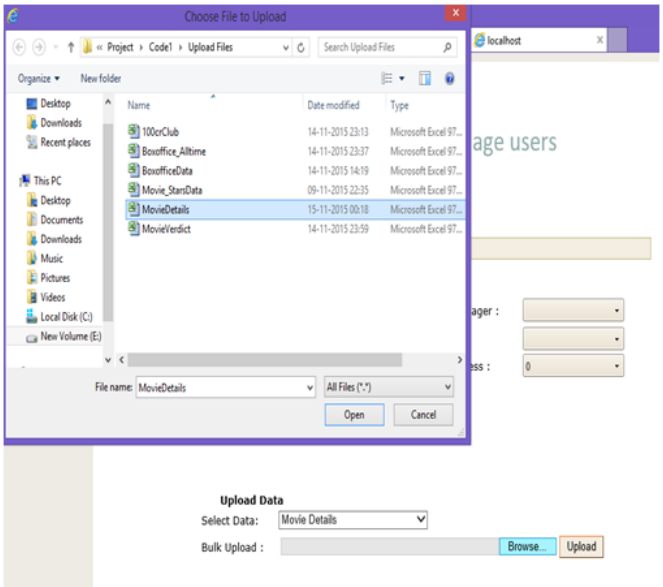


Figure 6.10: File selection for uploadng

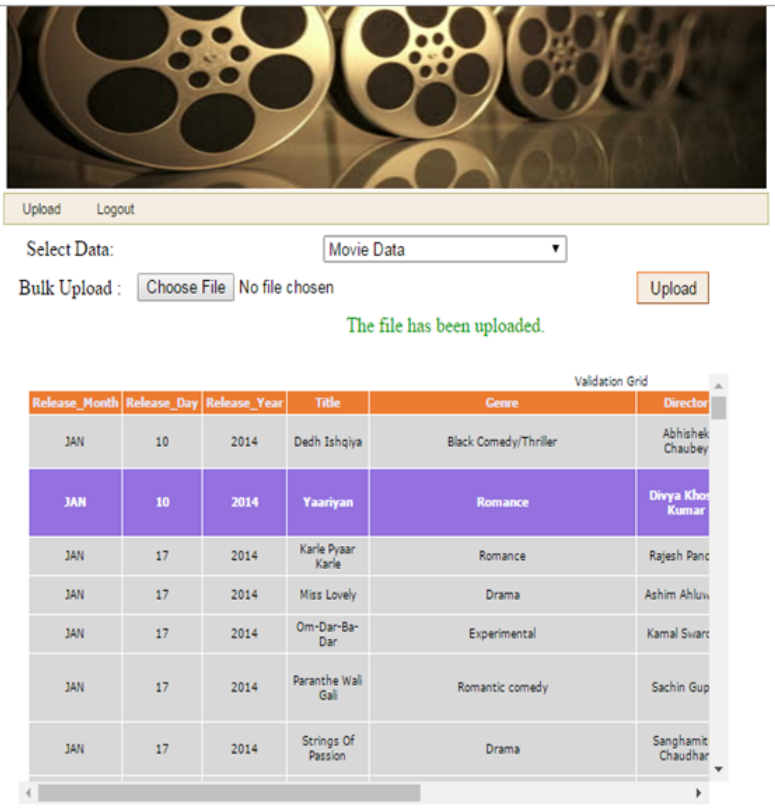


Figure 6.11: After upload

C. To upload data from front-end the admin has to enter the data in one of the 6 excel files available. Admin has to enter data in the format specified in excel file.

D. After selecting the required file admin has to click on upload button for data to get uploaded in database.

v. To get the forecasted verdict user has to enter the details as mentioned in the form below

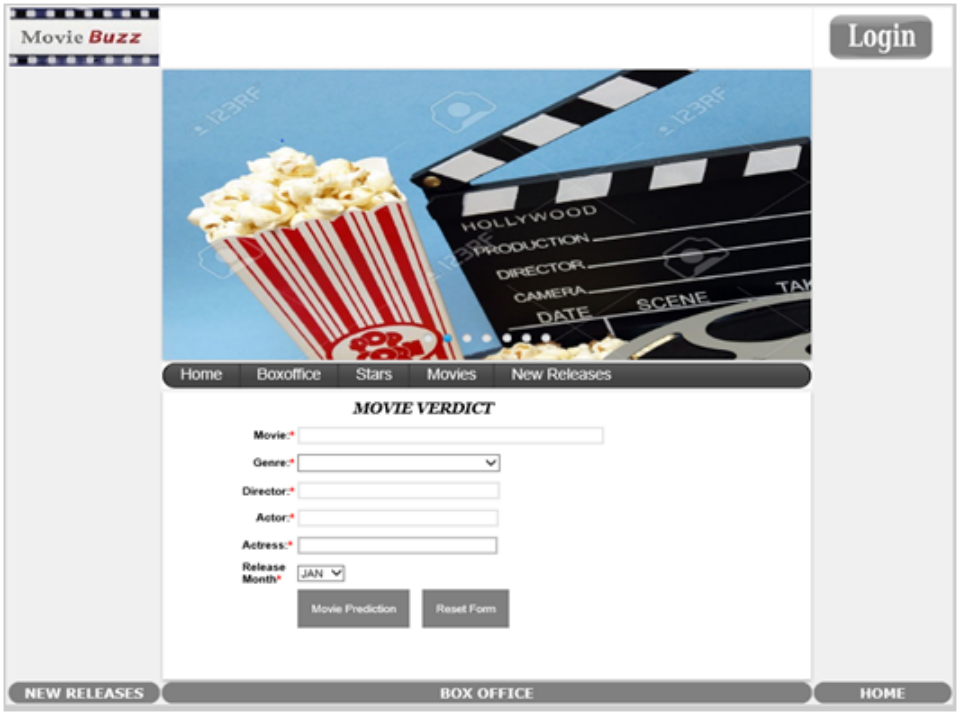


Figure 6.12: Input parameters

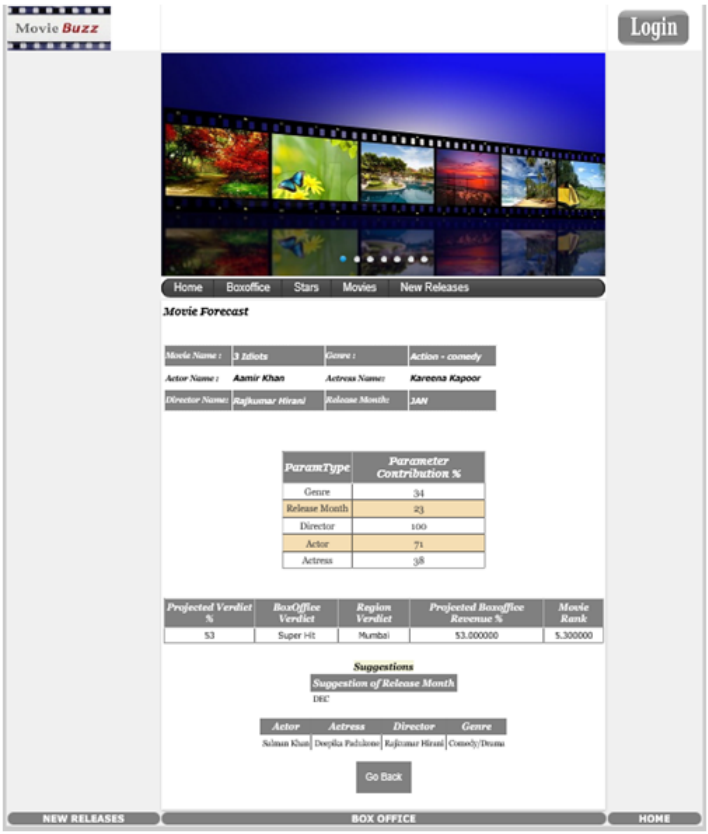


Figure 6.13: After entering the details the output shown is

6.2.3 Algorithms Used

Min-Max Normalization:
Normalization is done to scale the attribute data so that it fits in a specific range. There are many types of normalization techniques available like Z-score, Min-Max, etc. A simple Min-Max algorithm is used to normalize the data since using it we could normalize our values in the range of 0 to 1.

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} * (B - A) + A$$

Y is the transformed normalized value and X is the value which you want to normalize

If you want to convert the value in to range [0, 1] then A is 0 and B is 1.

Regression Algorithms:

1. Linear regression algorithm - Linear regression involves a response variable y and a single predictor variable x. For linear regression value of y will be approximately calculated as follows:

   $$y = Ax + B$$

   Here A is the slope and B is the intercept.
   Many prediction projects use Linear regression algorithm for predicting results [3].But since our verdict is depended on many variables we used Multiple linear regression algorithm to forecast.

2. Multiple-Linear regression algorithm - It is used when there is 1 response variable y and many predictor variables for e.g. x1, x2, x3, x4... Here the function will look like

   $$y = A0 + A1xi1 + A2xi2 + ... + Anxin + \epsilon$$

for i=1,2, ...k. where A0 is the intercept. A1, A2, ... , An are partial regression coefficients and $\epsilon istherandomerror$

6.2.4 Pseudo Code

Calculations of outputs are as follows

1. Verdict Calculation:
   (a) Actor
   (b) Actress
   (c) Genre
   (d) Director

(e) Release Month

Contribution of each parameter to calculate the verdict is as follows: For e.g. consider the Genre entered as Drama

(a) Then the movies net gross is selected whose genre is as mentioned above.

(b) Data is normalised year wise

(c) Net grosss is normalised by using the following formula v'=((v-min)(newmax-newmin))/(max-min)+newmin

Where v=the gross which is to be normalised
max=max of that year
min=min of that year
new max=1
new min=0 and v=normalised value
After the data is normalised the contribution of genre for success of that movie is as follows:

(d) From the normalised data maximum value is selected

(e) Then each value is divided by maximum and then average of all the values obtained is calculated. In the same manner the contribution of other parameters is calculated.
Projected verdict= Average of (genre, actor, actress, release month, director)
Verdict= If Projected verdict less than 35 Verdict=Flop
If Projected verdict greater than equal to35 and less than 43 then Verdict=Semi hit
If Projected verdict greater than equal to 43 and less than 65 then Verdict=Super hit
If Projected verdict greater than equal to 65 and less than equal to 100 then Verdict=Block Buster

2. Region Where Movie will do good business calculation

   (a) Actor ,Actress and Genre Parameter entered above are considered here

   (b) Net gross of movies is considered which consist of actor, actress or genre i.e.parameter which has highest number of movies is considered

   (c) The gross of the movies available is divided by the highest gross available in that table.

   (d) After division the values are grouped region wise. Average of all the movies belonging to same region is done .The region which has maximum average is considered as the region where the movie may do good business.

3. Movie Rating= The projected verdict calculated in output 1 divided by 10.

4. Suggestions for the movie. The releases month suggested is as follows:

   (a) Actor, Actress and genre Parameter is considered.

   (b) Net gross of movies consisting of actor or actress or genre is considered.

   (c) Netgross is divided by maximum net gross available in that table.

   (d) Average of divided value is calculated, which is grouped by months.

   (e) The month which gives the highest value is suggested as the month of release to increase the overall performance of the movie.

5. Estimated gross prediction It is calculated as follows
   Suppose the parameters entered above contributed as
   Genre = 32
   Release month=42
   Director=40
   Actor=41
   Actress=57

   Estimated net gross=(Average(32,42,40,41,57)*max(32,42,40,41,57))/100=23.94

Chapter 7

Software Test Document

This chapter discusses about the Software Test Document which describes the plan for testing the developed software system against the software requirements as defined in the Software Requirement Specification. The purpose of these system tests is to make sure that the software system developed complies with the definition of the software requirements. It describes the basic tests approach undertaken along with the Test Plan which will include the following: features to be tested, testing tools and environment. It also includes the test specifications.

7.1 Introduction

7.1.1 System Overview

Movie business in India is a billion dollar industry, employing thousands of people, as such success or failure of a movie can have profound effect on the stake holders. It is therefore of prime interest of the stake holders to know how the movie will fare. This however, is not an easy work, since movies have been described as experience goods with very less shelf life; it is difficult to forecast demand for a movie. There are number of parameters that may influence success of a movie like time of its release, marketing gimmicks, lead actor, lead actress, director being some of the factors. Our system aims to develop a model that may help in forecasting the success of a movie in terms of - will the movie be hit flop or neutral, rank of

the movie, its box office revenue prediction, regions where the movie will do good business, suggestion for the movies release date to increase profit thereby reducing certain level of uncertainty.

7.1.2 Test Approach

Test	Description
Functionality Testing	All the links in web pages, database connection, forms used in the web pages for submitting or getting information from user will be tested.
Usability testing	The system will be checked to see whether the web site is easy to use, clear instructions are provided, main menu is provided on each page. The content on each page is checked.
Interface testing	The system will be tested to see whether all the interactions between the servers(web, application and database) are executed properly.
Compatibility testing	The system will be run on different web browsers to check its compatibility
Performance testing	The system will be tested for its heavy load sustainability. This includes web load testing and stress testing.

7.2 Introduction

7.2.1 Features to be Tested

The software features which are to be tested include:

1. Content in the Excel sheet
2. Data Pre-processing

3. Database loading through SQL server
4. Admin login
5. Uploading of data through GUI by admin
6. Table selection
7. Data in SQL server
8. Script for uploading data in the database
9. Forecasting script
10. Handling of input

7.2.2 Features not to be Tested

1. **Data extraction**
   Data extraction will not be tested, since data will be extracted from trusted sites like Wikipedia, Box-Office India and koimoi.com.

7.2.3 Testing Tools and Environment

1. Microsoft Excel
2. Microsoft SQL Server 2012

7.3 Test Cases

7.3.1 Data pre-processing testing TC1

7.3.1.1 Purpose

To check whether the processed data conforms to the data format required to upload it to the database.

7.3.1.2 Inputs

Raw data extracted from websites like Wikipedia, Box-Office India and Koimoi.com copied in the excel workbook.

7.3.1.3 Expected Output and Pass/Fail Criteria

The data in the excel workbook should conform to the format required for uploading it to the database.
Criteria-
Pass - If the data in the excel workbook conforms to the format required for uploading it to the database
Fail - If the data in the excel workbook has errors or it does not conform to the format required for uploading it to the database

7.3.1.4 Test Procedure

Data collected is pre-processed by using the command button added in excel workbook (i.e. macro) which is coded using VBA programming. When the user clicks the command button following options will be provided for pre-processing.

1. The user has to enter the month in the specified format.
2. User can delete rows if required
3. Delete column
4. Column can be made blank.
5. Data entry
6. Rank entry
7. Set date format
8. Enter dates.

The user has to select from the options available to pre-process thedata.

7.3.1.5 Actual Output

Data in the excel workbook is free from errors and is of the format required for uploading it to the database.

7.3.2 Testing admin login TC2

7.3.2.1 Purpose

To test the admin login authentication.

7.3.2.2 Inputs

1. Username
2. Password

7.3.2.3 Expected Output and Pass/Fail Criteria

The admin should enter username and password this entered username and password will be mapped with the username and password pair stored in the database. If there is a proper match then there will be successful login and the admin will be redirected to the upload page else an error message will be shown for the unsuccessful login attempt.
Criteria -
Pass - If the admin enters correct username and password and is redirectedto the upload page or if user enters wrong username or password and error message is displayed
Fail - If the admin enters wrong username or password and error message is not shown.

7.3.2.4 Test Procedure

Functional and Interface Testing will be employed for this test case.

1. Admin will enter the username and password
2. The admin will then click the Login button

7.3.2.5 Actual Output

Admin logged in successfully and was re-directed to admin interface or ifusername or password is entered wrong then appropriate error message should be displayed..

7.3.3 Testing uploading of data through GUI by admin TC3

7.3.3.1 Purpose

To check whether admin is allowed to upload the data using web application.

7.3.3.2 Inputs

The excel sheet with pre-processed data which needs to be uploaded.

1. Table name (Dropdown menu will be provided to select the table).
2. Path to reach the required excel sheet stored on the computer.
3. Clicking the upload button.

7.3.3.3 Expected Output and Pass/Fail Criteria

After following the test procedure for this test case the data in the excel sheet should be stored properly in the SQL server database without any errors and successful upload message should be displayed.
Criteria -
Pass - If the excel sheet is successfully uploaded to the SQL server database.
Fail If there are some errors while uploading the excel sheet to the SQL server database.

7.3.3.4 Test Procedure

A web application is created using visual studio where admin needs to login successfully after which he/she will be directed to the upload page. The admin must select the table in which he wants to upload the data and then browse for the excel sheet with data which he wants to upload and click the upload button.

7.3.3.5 Actual Output

Data in the excel sheet was successfully uploaded to the SQL server database and successful upload message was displayed.

7.3.4 Testing table selection TC4

7.3.4.1 Purpose

To test whether the admin is allowed to select the proper table where he wants to insert/upload the data and to test whether the data is uploaded successfully to the selected table.

7.3.4.2 Inputs

Dropdown menu containing names of the table - Movie verdict, Star data, Box-office data, Box-office all time, 100cr club, etc.

7.3.4.3 Expected Output and Pass/Fail Criteria

Admin is allowed to select the table in which he wants to enter the dataand uploading of data to that table takes place successfully without any errors.
Criteria -
Pass -If selection of the required table is allowed and data is uploaded tothat particular table.
Fail -If selection is not allowed or there are some errors with uploadingthe data to that particular table.

7.3.4.4 Test Procedure

Dropdown menu containing names of the table -Movie verdict, Star data, Box-office data, Box-office all time, 100cr club, etc. will be provided to the admin. Admin must select the table to which he wants to upload the data, give the path to the excel sheet, click the uploadbutton, check whether the data is uploaded properly and completely to that selected table.

7.3.4.5 Actual Output

Admin could select the required table and data was uploaded successfully to the selected table in the database.

7.3.5 Testing script for uploading data in the database TC5

7.3.5.1 Purpose

To check proper execution of the script created for uploading data in the database.

7.3.5.2 Inputs

Script for uploading data in the database

7.3.5.3 Expected Output and Pass/Fail Criteria

The script should be executed without any errors and data in the excel sheets should be uploaded to proper table created in the database.
Criteria -
Pass -If the script executes without any error.
Fail If there are errors while executing the script.

7.3.5.4 Test Procedure

This script is created to upload the data in the database through front end i.e. through web application. Interface testing will be employed forthis test case.

1. The script will be compiled.
2. The script will be executed.

7.3.5.5 Actual Output

Script was successfully compiled and executed. The selected data was uploaded in the database.

7.3.6 Testing forecasting script(Movie verdict script) TC6

7.3.6.1 Purpose

To check proper execution of the script created for forecasting results depending on users input and the data in the database.

7.3.6.2 Inputs

Script for forecasting calculating the following -

1. Projected contribution of input parameters
2. Forecasted verdict
3. Region where the movie will do good business.
4. Box office revenue prediction
5. Movie rating

   Writing query for execution of the forecasting script giving different inputs for each input parameter Genre, Release month, Director, Actor and Actress.

7.3.6.3 Expected Output and Pass/Fail Criteria

Script should forecast desired output with very less error or no error.
Criteria -
Pass Script is compiled and executed successfully and is able to forecastrequired output.
Fail Script is unable to forecast the required output or gives errors whilecompiling or executing

7.3.6.4 Test Procedure

As mentioned above in this test case various inputs will be taken from theuser and the script will be tested. Interface testing will be employed forthis test case.

1. The script will be compiled.
2. The script will be executed.

7.3.6.5 Actual Output

Script was successfully compiled,executed and required outputs are displayed.

7.3.7 Testing forecasting script(Release month suggestion script) TC7

7.3.7.1 Purpose

To check proper execution of the script created for forecasting results depending on users input and the data in the database.

7.3.7.2 Inputs

Script for forecasting calculating suggestion for release month. Writing query for execution of the forecasting script giving different inputs for each input parameter Genre, Release month, Director, Actor and Actress.

7.3.7.3 Expected Output and Pass/Fail Criteria

Script should forecast the release month.

Criteria -

Pass Script is compiled and executed successfully and is able to forecast release month.

Fail -Script is unable to forecast the release month or gives errors while compiling or executing.

7.3.7.4 Test Procedure

As mentioned above in this test case various inputs will be taken from the user and the script will be tested. Interface testing will be employed for this test case.

1. The script will be compiled
2. The script will be executed.

7.3.7.5 Actual Output

Script was successfully compiled, executed and required output is displayed.

7.3.8 Testing movie verdict page TC8

7.3.8.1 Purpose

To check outputs are shown in proper format to the users.

7.3.8.2 Inputs

User should enter the following:

1. Movie name
2. Genre
3. Release month
4. Director
5. Actor
6. Actress

And Click movie prediction button
Error message will be shown if any input field is empty.

7.3.8.3 Expected Output and Pass/Fail Criteria

All input parameters are entered and redirected to Movie forecast page where following forecasted outputs should be shown:

1. Projected contribution of input parameters
2. Forecasted verdict
3. Region where the movie will do good business.
4. Box office revenue prediction
5. Movie rating

Criteria -
Pass - If on clicking the movie prediction button redirection to movie forecast page should take place and all the above mentioned outputs should be displayed.
Fail - If on clicking the movie prediction button redirection to movie forecast page does not take place or all the above mentioned outputs are not displayed.

7.3.8.4 Test Procedure

User should enter all the mandatory input fields and click the movie prediction button. After this he will be redirected to movie forecast page and the above mentioned forecasted results will be displayed here.

7.3.8.5 Actual Output

User enters all the input fields and is redirected to the movie forecast page where all the forecasted outputs are displayed.

7.3.9 Testing of reset form button TC9

7.3.9.1 Purpose

To check if the form resets on clicking the reset button.

7.3.9.2 Inputs

User should enter the following:

1. Movie name
2. Genre
3. Release month
4. Director
5. Actor
6. Actress

   And Click reset button

7.3.9.3 Expected Output and Pass/Fail Criteria

All input parameters are entered should be reset i.e. the form should be empty
Criteria -
Pass - If on clicking the reset button the form resets.
Fail If on clicking the reset button the form does not reset.

7.3.9.4 Test Procedure

User should enter all the mandatory input fields and click the reset button. After this the input form should reset for the user.

7.3.9.5 Actual Output

On clicking the reset button input form resets successfully.

7.3.10 Testing of Movie details page TC10

7.3.10.1 Purpose

To check if the form resets on clicking the reset button.

7.3.10.2 Inputs

User should enter the movie name and click search button

7.3.10.3 Expected Output and Pass/Fail Criteria

If Movie details are shown if the movie is present in the database or error is displayed if movie is not present.

Criteria -

Pass -If Movie details are shown if the movie is present in the database or error is displayed if movie is not present

Fail If Movie details are not shown even when movie is present in the database or error is not displayed if movie is not present.

7.3.10.4 Test Procedure

User should enter any movies name and click the search button. Movie details are shown if the movie is present in the database else error is displayed

7.3.10.5 Actual Output

Movie details are shown if the movie is present in the database else error is displayed.

Chapter 8

Result And Discussion

This chapter provides details of the work done on the project and its analysis. It presents the results that were achieved after implementation.

8.1 Results and Discussions

1. Verdicts related forecasting
   The system was tested against training data of movie verdicts. The training data was found to be approximately 86 % accurate for the verdicts been projected.

2. Region related forecasting
   As per our study been done most of the movies do good business in the region of Mumbai. So Mumbai is considered as the best region for earning gross.

3. Star Power

   Actors, Actresses and Directors influence the success of movies more as compared to genre and the month in which the movie is going to be released. This gives a hint that the potential of directors and star cast forms the base in deciding the success of movie.

	Film	Genre	Release_Month	Director	Actor	Actress
2	PK	21	26	100	71	39
3	Bang Bang	19	16	63	58	51
4	Singham Returns	19	18	64	38	38
5	Jai Ho	34	23	34	56	34
6	Humpty Sharma k	22	21	23	69	69
7	Gunday	19	21	80	47	29
8	Entertainment	15	18	21	33	94
9	Humshakals	19	20	65	29	23
10	Haider	11	16	84	54	53

Figure 8.1: Contribution of each parameter

Chapter 9

Conclusion and Scope for future work

This Article gives an overview of the Scope for Future Work for the given project.

9.1 Conclusions

Through this project, we have tried to design a low-cost , user friendly and interactive website that provides an insight of a released movies success. This website will be beneficial to producers and owners of film production studios in order to plan their future releases as well as audience who can decide which movie to watch depending upon its success in boxoffice.

9.2 Scope for future work

In this website application we have tried to incorporate all basic features which the target audience want. Apart from these features following features can be added to enhance the functionalities and appeal of the application

1. Age-wise suggestion of movies
2. Number of screens the movie should be allotted

References

[1] Supreme Motnam C.-C. Chan Kathy J. Liszka Krushikanth R. Apala, Merin Jose and Federico de Gregorio1. Prediction of movies box-office success using social media. *IEEE*, 2013.

[2] Songkuk Kim Seonghoon Moon, Suman Bae. Predicting the near-weekend ticket sales using web-based external factors and box-office data. *IEEE*, 2014.

[3] https://en.wikipedia.org/wiki/list_of_bollywood_films_of_2015.

[4] http://www.koimoi.com/box-office-bollywood-films-of-2013/.

[5] http://www.koimoi.com/box-office-india-bollywood-100-crore-club/.

[6] http://www.bollywoodhungama.com/box-office/top-grossers=.

Acknowledgement

We take this opportunity to express our profound gratitude and regards to our guide Prof. Deepti Patole for her exemplary guidance, monitoring and constant encouragement throughout the course of this project.
We also take this opportunity to express a deep sense of gratitude to Head of the department, Prof. Sangeeta Nagpure for her cordial support, valuable information and guidance, which helped us in completing this task through various stages.
We are obliged to staff members of K. J. Somaiya College of Engineering, for the valuable information provided by them in their respective fields. We are grateful for their cooperation during the period of our assignment.

Date:

Sanjana Anaokar
Ria Marwaha
Pooja Karkera
Chaitali Tanawade