WEB APPLICATION FOR MOVIE PERFORMANCE PREDICTION

Devalkumar Patel

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd

Part of the Computer and Systems Architecture Commons

Recommended Citation
Patel, Devalkumar, "WEB APPLICATION FOR MOVIE PERFORMANCE PREDICTION" (2020). Electronic Theses, Projects, and Dissertations. 980.
https://scholarworks.lib.csusb.edu/etd/980

This Project is brought to you for free and open access by the Office of Graduate Studies at CSUSB ScholarWorks. It has been accepted for inclusion in Electronic Theses, Projects, and Dissertations by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
WEB APPLICATION FOR
MOVIE PERFORMANCE PREDICTION

A Project
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
Of The Requirements For The Degree
Master of Science
In
Computer Science

by
Devalkumar Patel

March 2020
WEB APPLICATION FOR
MOVIE PERFORMANCE PREDICTION

A Project
Presented to the
Faculty of
California State University,
San Bernardino

by
Devalkumar Patel

March 2020

Approved By:

Dr. Ernesto Gomez, Advisor, Computer Science and Engineering
Dr. Yan Zhang, Committee Member
Dr. Qingquan Sun, Committee Member
ABSTRACT

This is an amazing and unique idea for the web application. The purpose of this application is to address those movie lover people which is always in a hurry to visit the theatre to watch upcoming movies irrespective of which star cast is in it. This system is a mixture of desktop applications, python libraries, and simple math arithmetic. The application can be used by anyone which is ultimately helping them to decide, either they should watch the movie or not. The user just submits the name of movie. This application is developed in Visual Studio 2019 for functioning back-end and front-end while data cleaning and import have been done in Spyder IDE with 95% prediction accuracy.
ACKNOWLEDGEMENTS

I want to thank my committee chair and project advisor Dr. Ernesto Gomez who was continuously encouraging me for development and giving input for the project production. I want to express my gratitude to Dr. Yan Zhang who was continuously giving me new standards for the project and I learned a lot from them. I want to thank Dr. Qingquan Sun who has given me the time from his busy schedule. It is my pleasure that I got such fantastic committee members.

I want to thank my parents who have continuously supported me for education financially, ethically and morally.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td>2</td>
</tr>
<tr>
<td>Existing System</td>
<td>2</td>
</tr>
<tr>
<td>CHAPTER TWO: SYSTEM ANALYSIS</td>
<td>4</td>
</tr>
<tr>
<td>Proposed System</td>
<td>4</td>
</tr>
<tr>
<td>System Requirement Specification</td>
<td>5</td>
</tr>
<tr>
<td>CHAPTER THREE: FEASIBILITY REPORT</td>
<td>6</td>
</tr>
<tr>
<td>Feasibility Study</td>
<td>6</td>
</tr>
<tr>
<td>Technical Feasibility</td>
<td>7</td>
</tr>
<tr>
<td>Economic Feasibility</td>
<td>7</td>
</tr>
<tr>
<td>Operational Feasibility</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER FOUR: SYSTEM DESIGN</td>
<td>10</td>
</tr>
<tr>
<td>UML Diagram</td>
<td>10</td>
</tr>
<tr>
<td>Use Case Diagram</td>
<td>10</td>
</tr>
<tr>
<td>Sequence Diagram</td>
<td>12</td>
</tr>
<tr>
<td>Activity Diagram</td>
<td>13</td>
</tr>
<tr>
<td>Data Flow Diagram</td>
<td>15</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1: Use Case Diagram.................................................................................................11
Figure 2: Sequence Diagram.................................................................................................12
Figure 3: Activity Diagram..................................................................................................14
Figure 4: DFD Symbols........................................................................................................16
Figure 5: Level 0 Data Flow Diagram..................................................................................17
Figure 6: Level 1 Data Flow Diagram................................................................................18
Figure 7: Level 2 Data Flow Diagram................................................................................19
Figure 8: Starting Page Of Application................................................................................29
Figure 9: Entering correct Username And Password .......................................................30
Figure 10: Successfully Authenticated................................................................................31
Figure 11: Entering Wrong Credentials............................................................................32
Figure 12: Not Authenticated..............................................................................................33
Figure 13: New Movie Entry Page......................................................................................34
Figure 14: New Film Crew Entry.........................................................................................35
Figure 15: Connecting To The Database............................................................................36
Figure 16: Back-End Performance Indices.........................................................................37
Figure 17: Predicted Result.................................................................................................38
Figure 18: Actual Result.......................................................................................................39
Figure 19: Predicted Result.................................................................................................41
Figure 20: Real-Time Result...............................................................................................42
CHAPTER ONE

INTRODUCTION

Background

The box office industry is one of the largest industries in the world that is catering to billions of jobs in the world. There are different movie industries that have flourished around the world depending on the countries, languages, and province. As time goes on, these industries have evolved. More technology has been added in these industries like cameras, editing, special effects, graphics, distribution, and more. Each industry is thinking about the best performance for what they want to remove the hurdle and want to propel their business. In the movie industry, the biggest issue is a lack of trust in upcoming movies.

The industry’s promotion and marketing teams have addressed this issue by launching trailer, teaser, posting advertisements on television, magazines and social media platforms, hosting a social event like a fundraiser or a donation for a cause. Despite of various campaigns by the marketing and publicity teams, good movies do not perform well as they deserve. My project aims to strengthen the trust in viewers by giving prediction whether a movie is a hit or a flop so a person can decide to watch the movie or not. This application currently works on the Indian Bollywood industry due to small data size and quality of data.
Purpose

The purpose of this application is to provide the prediction of performance for an upcoming movie in theatres by calculating the previous movie performances of the star cast, director, producer, and music director from historical datasets using a simple arithmetic equation.

This application takes account of a member from lead star cast to the music director. A movie will perform well when every member of the team works in cohesion. This application has a 95% accuracy for a prediction from the available small dataset. Based on the predictions people can decide to watch the movie or not and can save money. If a person could decide by their own self then the marketing and publicity teams could be less stressed about PR management, marketing, and all activities for reaching out to audiences.

Existing System

In the current world, there is no such application exist which can predict a movie’s performance. Yes, there are many critics who give their reviews, and platforms where a person can go and submit their personal feelings. These all are made possible after the movie is released. The people who want to watch the movie remain perplexed by reading reviews published on various platforms.
Right now, the focus is on the lead actor and actress, marketing activities like selling merchandise, print media management, promotional activities and many more. Due to all these promotional activities a hype is created for the movie, people fall in trap of these promotions and join the hype.
CHAPTER TWO

SYSTEM ANALYSIS

Proposed System

This application works on the average function which runs in two stages. In first step, average function calculates the rating on each crew member which is stored in the dataset, the output generated in first stage is given as an input to the second stage. In the final stage the application gives the verdict whether movie is hit or flop. Function is based on the average calculated on the team members and a result is generated. The application can compile and run after satisfying the below requirements. The version of C# should be greater than 16 and visual studio 2019. These requirements specified are the minimum, any version below will not support the application. Application is hosted on a real-time Wamp server for executing any request. A Wamp server can be hosted on mobile or desktop. If a user demands the result of a movie and the data for the movie is not available in the historical dataset for some crew members. then, the user can add the rating for that missing crew member. Yet, if data is not available then it will not affect the output generated. After the result is produced it is saved in the database once for each movie and respective crew members. Over time dynamics in data keeps training the system.
System Requirement Specification

This project consists of four phases data importing, cleaning, development and production. The prototype’s minimum hardware and software requirements are as follows.

❖ **Hardware Requirement:-**

- i3 Processor Based Computer
- 1 GB-Ram
- 256 GB Hard Disk

❖ **Software Requirement:**

- Windows XP, Windows 7(ultimate & enterprise)
- SQL Server 2008.
Feasibility Study

Feasibility Study is the most important part for the development. It studies aspects of the present, experience and future and tries to answers that would arise in any project manager’s mind. The feasibility study is conducted when the problem is answered and understood. The feasibility study will determine if a proposed system deserves an investment of time and money, if the idea is implemented will it solve any proposed issues or not? Through the feasibility study, the management can visualize the proposed system. In the study, all the attempts by others will be considered.

The most popular decision-making feasibility studies are Technical, economical and operational. The feasibility study considers product idea, on-ground demand, technical feasibility and many more. when report is generated all the future problem will be address and eliminated them before they arise.

Here, we are going to study three different feasibility aspects for the proposed-system which are technical aspects, economical aspect, and operational aspect. These aspects also addresses the new proposed system.
Technical Aspect

Technical feasibility is providing the estimation for a minimum amount of technical skills required. Upon the estimation, a developer can predict whether they must gain knowledge or not. Which will ultimately affect the time and economic feasibility? This is a cascading process. In this estimation following are the requirements.

➢ This system is maintainable for new updates or release
➢ This system should be secure, scalable and optimum.
➢ This system is responsive.

When all the requirements are met the system is said to be technically feasible

Operating System : 7(ultimate & enterprise)
Languages : Asp.Net with C# (.Net 2010)
Database System : MS-SQL Server 2008
Documentation Tool : MS - Word 2013

Economic Feasibility

Economic feasibility is one of the major factors. In this project one say that this project is financially feasible because it requires no extra financial investment, this project can be completed in six months.

In economic feasibility, existing projects are compared based on future financial benefits of the proposed system with the equity, return on investment is
calculated. Profits are expected on the investments. If a system gives more returns compared to investments then the system is financially feasible.

Economic Feasibility will conclude whether the project can be done from available resource or not. Financial benefits should be greater then investment. In this feasibility following are checked:

- Cost price for full system investigation.
- The installation and operational cost for the system.
- The cost during the development includes tools.

The returns of this project comparing to the investments is very good. So this project is economically feasible because it is giving more returns compare to investments.

**Operational Feasibility**

In this step, all different operational factors are determined of the proposed systems like work force, time, etc. to design, develop and maintain this system. whichever solution uses least amount of effort, is the best operationally feasible solution. Operational Feasibility determines whether the proposed system satisfies all clients' demands and functionalities.

- Work culture, style of executions, planning and the design of the product should be accepted by all the members. The proposed system will not cause any problem under any circumstances to the end-user and works properly.
In this feasibility study the time and personal requirements are satisfied.

so this solution is operationally feasible
CHAPTER FOUR
SYSTEM DESIGN

UML Diagrams

Use Case Diagram

This is the most integral part as one can understand the system logically on over all the scenarios of the system. this part determines all the requirements which would influenced the system internally and externally. This presents the way how a user interacts with the system. So when the requirements are drawn, we analyze the system, the use cases are generated and entities are identified. It provides all available ways where user interacts with the system. the regular way of developing a use case is to break the system into small parts and develop separate use cases In this way we eliminate replication In the diagram. There are multiple roles that users have to perform in the system environment. The most feasible solution is by providing the use case diagram, it rules out the confusion from the user. This diagram is updated whenever any changes happen in the system.

The use case diagram shows the functionality of the system which has been provided. In use case diagram language end-user is called as an actor. In this system, there has only one actor and that is admin.
Figure 1: Use Case Diagram

Admin

- Access the Website
- Login as Admin
- Checks for new movie
- Adds the Movies
- Adds the History of Movies & Directors
- Rates a particular movie
- Manages all Security Related issues
Sequence Diagram

Figure 2: Sequence Diagram
Activity Diagram

START

Admin

Login

Invalid Login

If Valid

Access to Main Page

Checks for New Movie

Adds New Movies

Adds History of Movies & Directors

Rates a particular movie

Manages all Security Related issues

STOP
Figure 3: Activity Diagram
Data Flow Diagram

Data flow diagram is an integral part of documentation. When a developer is stuck this diagram is referred. A data flow diagram gives the visualization of data movements in the system from one part to another. It provides information about manipulation and processing of the data. hence it called logical data flow diagram. Diagram consist of three levels from level 0,1 and 2. The process in each level is the base for the next level. The upper-level diagram is called a context-free diagram which is a key player for studying the current system. the exploration of the level continues until the analyst perspective is met.

Data Flow Diagram is called a bubble chart which is exposing the requirements for the system and major transformation for the system design. there are many symbols for diagram.
Figure 4. DFD Symbols

- **Process that transforms data flow.**

- **Source or Destination of data**

- **Data flow**

- **Data Store**
Figure 5. Level 0 Data Flow diagram
Figure 6. Level 1 Data Flow Diagram
Figure 7. Level 2 Data Flow Diagram
CHAPTER FIVE

SYSTEM TESTING

Testing

The testing phase is a part of system development lifecycle. Testing phase gives confidence to the developer in the system. During testing phase, different methodologies are considered like Unit testing, integration testing, and system testing. Testing phase determines whether application meets the requirements and serves the functionalities. This system is developed in C# for the web application.

For large scale project these tests has to be successful. If each part of the system is working correctly and is behaving or providing the desired output for every kind of input then this system or project is successfully tested. Here success means system has perfect data flow, scalability, availability for updates and reacts perfect on each input. Successful testing means the system ready to deploy. in the documentation language testing phase is important as It is the foundation of our whole project for the current and future.

There are different phases in testing:

1) Unit Testing
2) Integration Testing

3) System Testing

4) Validation Testing

5) Output Testing

6) User Acceptance Testing

**Unit Testing:**

This is the first phase of the testing. It provides the inner view of the project if there has any defect. Basically, in this phase, we are dividing the system into a very small working unit. We are testing that small unit of code. So, we can get to know if there was any error at the basic implementation level and can fix that. Majorly developer during development of that module has been tested. So, there has some time slice for testing during the development phase.

**Integration Testing:**

As the name suggested, during integration testing we have to integrate something and form like a structure. We have to tests that structure whether it is working perfectly or not. Here is testing language, that something part is nothing but the separate module of the code. We are integrating that module and form like a small structure that is working as a part of the system.
In the unit testing, we have checked that each module is working correctly and there has not any error. But, sometimes individually working module is not giving perfect output due to the integration. They can not orchestrate with each other. So, integration testing is providing a clear view of the small structure. During testing, the system interface also tests with requirements.

**System Testing:**

System testing is the third part of the testing life cycle. Up to this stage we have tested our project as an individual and tested as a small part of the system. Now, in this phase, we have to test whether the whole system will be work properly or not. This phase is not only limited to the system only but it is also checking about the system that how it is successful to adopt new changes, testing about scalability, compatibility, acceptability of new updates and many things. In short, in this phase, we are checking our system on the performance aspect as an end-user. We can say that if each section of the project is working peacefully then this phase of testing is successful. Means our system has passed the system test.
**Validation Testing:**

This is the final section of the testing life cycle. In this section, we are majorly checking about the performance of the system that whether the system performance meets the requirements. After the testing of each module, integrate them and test that structure and after analyzing the system nature this phase has come.

This phase of testing of the out system is not creating any kind list that still needs to address. So, we can say that our system has been successfully tested with validation testing.

**Output Testing:**

This is the second phase of the final testing. This phase has come after the validation testing and no system will be useful if it is not providing the specified output that should be. The format requirement of the users' test by the generated output of the proposed system. the output format is considered in two ways one is on the screen and the other is the printed format and they both are matching with each other. So we can say that the system has successfully passed the output testing.
User Acceptance Testing:

This is the final and last part of testing. Here we are including the audience who is going to use this system as a product. They have been continuously involved and monitored if there have any changes provided by the end-user which called updates. This is the phase where the system gets tested.
CHAPTER SIX
DATA IMPORT FROM REST API

There are several ways to import data to your system. You can import data from local storage, integrate data from different services and combine them together or requesting data to the server which is not part of your environment with permission and your system and web server will exchange the data in JSON format.

1). Importing data from the local system:- sometimes you have to import data from local storage to your tool in the format of Excel, SAS,.CSV and many more. You can import data by reading the file and store it to the data frame.

2). Importing data from a web server (REST APIs):- Now, there may be a chance that data stored on local devices is not sufficient. you require real-time(Twitter streaming), more amount of data(Facebook, Google Analytics and many more) or frequently changing data (Stock market.) so you have to connect with the server and sync that data to your system. To fetch the data from the server you have to communicate with the REST APIs.

REST API is the abbreviation of (Representational State Transfer Application Programming Interface). That is a server that you can use to retrieve
data. You are requesting the API and getting the response in terms of data. To request data to API python using the Requests library. And in that, you are using Get protocol. There have 4 different protocols in the REST API. Get, Put, Post and Delete. For fetching data from the server we are using Get protocol. This data communication between the system and the server happens in the JSON data format. You are getting every property in response from the server like header, body, status_code and many things. So, you can access that information. To read data from the response, you have to access the JSON method. Now, you have JSON data and convert it to the data frame using pandas libraries.

We have imported data from the BookMyShow.com server which is located in India for real-time Hindi movie ratings between the period of 2015-2019. We have followed all the methods which have described above. So, our projected system has a good amount of data to generate the result.

**ASSUMPTION:** we have a tiny data set which consists of 165 movie entries from 2015-2019. And we are going to use data for training from 2015 to 2018 (140 entries), and for testing or prediction, we are using the year of 2019 (25 entries).
Data Cleaning is the most important and rigorous part of the job. Which consuming much amount of energy of Data Engineer. When you get the data there have much erroneous information and incompatible data like missing information, NaN, putting wrong data type like column is only for boolean data type and data entry is like string and many more. So if you want to run your system very smoothly and dint want any kind of problem then you have to remove this inefficiency.

In data cleaning, there have two approach

**Replacing value:** Developer doesn’t want to keep the garbage (NaN, N/A, changing the value of the attribute, etc) value in the data set. So they are replacing the value with some productive value (real-time value). But there have chances of effecting over-all data which will lead to erroneous results. So the developer replacing that garbage value with the median value of data set majorly so that can’t affect the result. But, sometimes that is affecting the outliers. Means, if the real value is very far than replacing value or you may consider that it is residing in the outliers then may affect the result on a particular attribute.
**Cleaning attribute:** replacing is not the only option. You can just put the null value for that attribute also. This is one of the best ways to clean data. You can just symmetrize the null value like replacing NaN or N/A values with NA. This approach is good particularly for the outliers because they are not replacing. Putting the null value in the cell is also not affecting the result. To extracting and manipulating the cell values you can use pandas and NumPy library. Convert that data set into a data frame and accessing column and each value. For taking whole values of the attribute then you can use NumPy array to access the value.

In our proposed system, we have taken the second approach on demand. We are keeping null values in a cell they just now symmetrize using Pandas and NumPy libraries. For generating the final result, it is taking so many input values so ultimately that will not affect.
CHAPTER EIGHT

OUTPUT SCREENSHOT

Figure 8: Starting Page Of Application
Figure 9: Entering correct Username And Password
Figure 10: Successfully Authenticated.
Figure 11: Entering Wrong Credentials
Figure 12: Not Authenticated
Figure 13: New Movie Entry Page

Enter movie detail of upcoming movie
there may be a chance that information not available of particular crew members if they are coming after a certain year of break or they are doing debut. If you don’t have a record of crew member then you have a choice to enter the detail of crew member of the movie
Figure 15: Connecting To The Database

It is Fetching values from the database and performing the calculation on that
Figure 16: Back-End Performance Indices.
Figure 17: Predicted Result

Entering the value of movie name and generating the result
This has shown in the market. It has only 4.6 stars. so it is below 5. And below 5-star rating movie is considering as a flop movie. This movie has been released in 2020, which is not in our trained data set. For this system, this movie is going to release in the future because we have data up to 2019. So, the result of this
system and the real-time result is matching and this system has predicted the perfect result.
Figure 19: Predicted Result

We are going to predict for another movie. Which has been released in 2020 which is again not in our dataset.
Figure 20: Real-Time Result

Predicted result is matching with the real time result
CHAPTER NINE

FUTURE ENHANCEMENT

Enhance Data Quantity

We can import more data before the year 2015. It will create more data pool and reduce the chances of availability of information. When we have more amount of data at that time we have more choices of training and testing of data. If your data is more trained then that will be more accurate.

Include Feature Of Artificial Intelligence

After having a bunch of data then you can trained model and getting the advantage of machine learning and artificial intelligence. That is the future of computing.

Adding NLP Feature

Here, we are getting results only on the bases of rating but we can't identify the feeling of the viewers. When we will have data of comments then we can do sentiment analysis and also prepare the data for a suitable genre for each crew member. Getting that data into the calculation that will create more accuracy.
CHAPTER TEN

CONCLUSION

This system has a very big scope to be useful in the real market and satisfy the requirement. As of now, this system has a 95% accuracy, using testing data set of 25 movies and using simple arithmetic calculation, which is amazing. There was a chance of may reduce the number when you include more data into the data set.

In conclusion, we can say that this system is producing the result which is matching real-time results generated by other web sites. So, this system can use as commercially and cater to the demand of users.
BIBLIOGRAPHY

1). Wikipedia: en.wikipedia.org/wiki/Arithmetic

2). Microsoft dotnet community: dotnet.microsoft.com/apps/asp.net This is able to clear all questions and doubts related to language


4). Tutorial: docs.python.org/3/tutorial/ this python tutorial is very good for beginner

5). Blog: www.dataquest.io/blog/python-api-tutorial/ This blog is full of usage of Python API in real time application.