

# Movie Recommendation System using Filtering Approach

Parth Kotak

Department of Computer Engineering  
Vidyalankar Institute of Technology  
Mumbai, India

Prem Kotak

Department of Computer Engineering  
Vidyalankar Institute of Technology  
Mumbai, India

**Abstract**— A recommendation engine filters the data using different algorithms and recommends the most relevant items to users. It first captures the past behavior of a customer and based on that, recommends products which the users might be likely to buy. If a completely new user visits an e-commerce site, that site will not have any past history of that user. Our project aims to implement a recommendation engine that responds to the user to get the recommendations for a movie. The ultimate purpose of Movie Recommendation System is to make the user's experience better by recommending them movies. In this we have performed Exploratory Data Analysis firstly and then we have created the Recommendation System and here we have created the correlation matrix to find top matches that relates the best for a particular movie after that we checked the result & saved it to a csv file. We have created our model by reusing the saved file to get recommendations and users have to search the movie name for a year and he/she will get the four recommendations. The System recommends the same movies to users with similar demographic features. Since each user is different, this approach is considered to be too simple. The basic idea behind this system is that movies that are more popular and critically acclaimed will have a higher probability of being liked by the average audience.

**Keywords**— Recommendation Services, Machine Learning, Exploratory Data analysis, Golden Model, Collaborative filtering

## I. INTRODUCTION

The project focuses on a recommendation engine that filters the data using different algorithms and recommends the most relevant items to users. It first captures the past behavior of a customer and based on that, recommends products which the users might be likely to buy. If a completely new user visits an e-commerce site, that site will not have any past history of that user. In this we have performed Exploratory Data Analysis firstly and then we have created the Recommendation System and here we have created the correlation matrix to find top matches that relates the best for a particular movie after that we checked the result & saved it to a csv file. We have created our model by reusing the saved file to get recommendations and users have to search the movie name for a year and he/she will get the four recommendations.

## II. LITERATURE SURVEY

### A. Content Based Filtering

Content-Based Filtering are also known as cognitive filtering. This filtering recommends item to the user based on his past experience. For example, if a user likes only action movies then the system predicts him only action movies similar to it which he has highly rated. The broader

explanation could be suppose the user likes only politics related content so the system suggests the websites, blogs or the news similar to that content. Unlike collaborative filtering, content-based filtering do not face new user problem. It does not have other user interaction in it. It only deals with particular user's interest. Content based filtering first checks the user preference and then suggest him with the movies or any other product to him. It only focus on single user's ideas, thoughts and give prediction based on his interest. So if we talk about movies, then the content based filtering technique checks the rating given by the user. The approach checks which movies are given high ratings by the user by checking the genre categories in the user profile. After analysing user profile, the technique recommends movies to user according to his taste.

### B. Collaborative Filtering

The concept of collaborative filtering was first introduced in 1991 by Goldberg et al. The Tapestry system applies only to smaller user groups (e.g. a single unit), and has too many demands on the user. As a prototype of collaborative filtering recommendation system, Tapestry presents a new recommendation, but there are many technical deficiencies. Since then, there has been a scoring based collaborative filtering recommendation system, such as Grouplens, which recommends news and films. At present many ecommerce sites have been using the recommendation system such as Amazon, CDNow, Drugstor and Moviefinder etc. There is massive amount of data available. As we all know that today in this busy life no one has time to search hundreds of thousands of item and select the one which is similar to their taste. So collaborative filtering is one of the ways to filter the data and provide the relevant information in which the user is interested in. Collaborative Filtering is one of the most well known techniques for recommending items. This technique suggests relevant item to the user based on neighbour's choice. It first finds out the similarity between the user and his neighbour and then predicts the items. There can be n number of users. This technique finds the similar user from the list of user's. But the similarity between users is found out based the ratings which the users have given to the particular item. This way the approach continues and the desired result is generated. This strategy takes ratings given by user for any item from the large catalog of item catalog of ratings given by the user.

### III. TECHNOLOGY USED

#### A. Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

#### B. Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

#### C. Google Colab

Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs

### IV. PROPOSED SYSTEM

#### A. Pre-processing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model.

We used the following Pre-processing components in our Project:

1. Raw Data- Raw data also called source data, atomic data or primary data is data that has not been processed for use.
2. Structured data- It is data that adheres to a predefined data model and is therefore straightforward to analyse.
3. Data preprocessing-It is a process of preparing the raw data and making it suitable for a machine learning model.
4. Exploration Data Analysis(EDA)-Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses and to check assumptions with the help of summary statistics and graphical representations.
5. Insight Reports, Visual Graph-Data insights refer to the understanding of a particular phenomenon you are able to achieve by using machine learning and artificial intelligence (AI) technology to analyze a dataset and presenting it in the form of Graphs.

#### B. Training Model

The process of training an ML model involves providing an ML algorithm, the learning algorithm with training data to learn from.

#### C. Prediction

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome.

### V. METHODOLOGY

Our project uses the Collaborative filtering which is provided by Goldberg et al.

The System uses the following methodology:

1. Raw Data
2. Pre-processing
3. Structured Data
4. Learning Algorithm
5. Candidate Model
6. Deploy Selected Model
7. Golden Model

### VI. ANALYSIS

#### Iterative Development Process Model

Iterative development model aims to develop a system through building small portions of all the features, across all components. We build a system which helps to analyse ratings of customers on basis of which it recommends movies to the user.

The phases of iterative development are:

#### 1. Planning:

As with most any development project, the first step is go through an initial planning stage to map out the specification documents, establish software or hardware requirements, and generally prepare for the upcoming stages of the cycle.

#### 2. Requirements:

In this phase, requirements are gathered from customers and check by an analyst whether requirements will fulfil or not. Analyst checks that need will achieve within budget or not. After all of this, the software team skips to the next phase.

#### 3. Design:

Once planning is complete, an analysis is performed to nail down the appropriate business logic, database models, and the like that will be required at this stage in the project. In the design phase, team design the software by the different diagrams like Data Flow diagram, activity diagram, class diagram, state transition diagram, etc.

#### 4. Implementation:

With the planning and analysis out of the way, the actual implementation and coding process can now begin. All planning, specification, and design docs up to this point are coded and implemented into this initial iteration of the project.

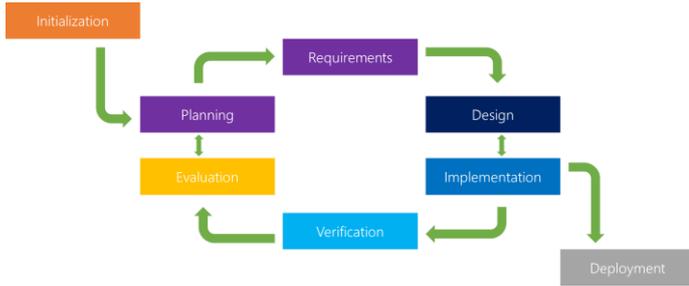
#### 5. Verification:

Once this current build iteration has been coded and implemented, the next step is to go through a series of testing procedures to identify and locate any potential bugs or issues that have cropped up.

#### 6. Evaluation:

Once all prior stages have been completed, it is time for a thorough evaluation of development up to this stage. This allows the entire team, as well as clients or other outside parties, to examine where the project is at, where it needs to be, what can or should change, and so on.

7. **Deployment:**  
 After completing all the phases, software is deployed to its work environment.



### VII. FEASIBILITY STUDY

Technology Considerations Movie recommendation systems available in the market are dependent on the dataset to contain large clusters of similar users and items. They also do not provide services such as effective remote access via cloud, customer interaction modules, etc. to be solved with the proposed system.

1. Product/ Service Market place The Movie recommendation system will impact client institutions in several ways. The following provides a high level explanation of how the organization, tools, processes, and roles and responsibilities will be affected as a result of the movie recommendation system implementation:-

2. Tools: The existing requirement for on site management systems will be eliminated completely with the availability of a cloud based system.
3. Processes: With the Movie recommendation system comes more efficient and streamlined administrative and customer relations processes. Hardware/Software: Clients will need to handle no extra software or hardware apart from a stable high speed Internet connection and a computer device.
4. Operational Feasibility The project will be implemented in a way that it will allow the functioning of recommendations smoothly. It will provide a user-friendly user interface in a modular fashion.

### VIII. COST ANALYSIS

**Infrastructure services**

These services include infrastructural components such as where the model is hosted, where data is stored and how the data is delivered. All these also need redundancies and load balancers for backup and security servers, which add both the cost and complexities.

**Servers -** Servers are where the app will be hosted. Some of the popular companies that provide hosting services are amazon aws, google gcp and azure.

We can use this to implement it globally over the internet. It has some the cost for training ML

GCP: \$0.54 per hour

Azure: \$9.99 per ML studio workspace per month \$1 per studio experimentation hour

AWS: \$0.42 per hour

If we need to increase the number of GPUs that your model supports, we will have to pay \$3.06 per Hour in AWS and

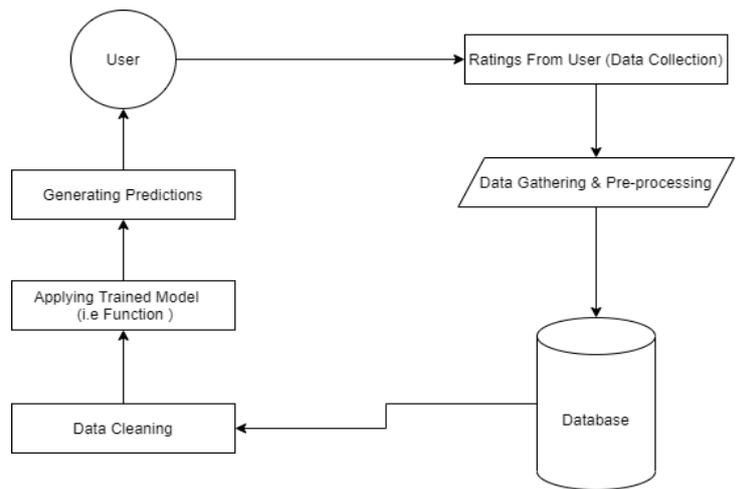
GCP Charges hourly with different cost depends on the GPUs.

ML model training – For the training of the AI - agent we will be using our local system with the following system specifications:

Component	Model	Cost
CPU	i7 8750h	₹ 28,991
GPU	GeForce gtx 1050 TI	₹ 26,850
RAM	16GB RAM	₹ 8,000

We are using the Jupyter Notebook and Google Colab for testing and checking the prediction of our model

### IX. DESIGN



### X. CONCLUSION

In this paper we have introduced and design a technique of filtering on a Data Base of Movies. It collects ratings from the user known as Data Collection and then pre-processes it. Furthermore, Data cleaning occurs followed by Training the ML model and then generating predictions. The User enters in the search bar the movie name and the year and gets recommended 4 movies depending on the likability and user ratings of a particular movie in that particular year. The models efficiency increases with a better dataset and prediction.

### XI. REFERENCES

- [1] [http://www.rieproject.com/article\\_106395\\_c6c0038f1bf5d4c421bd552d0541d6be.pdf](http://www.rieproject.com/article_106395_c6c0038f1bf5d4c421bd552d0541d6be.pdf)
- [2] <http://www.ijstr.org/final-print/dec2019/A-Review-Paper-On-Collaborative-Filtering-Based-Movie-Recommendation-System-.pdf>
- [3] <https://ieeexplore.ieee.org/document/8663822>
- [4] <https://www.mygreatlearning.com/blog/masterclass-on-movie-recommendation-system/>
- [5] <http://dx.doi.org/10.1007/978-981-13-1742-2>
- [6] <https://www.irjet.net/archives/V5/i3/IRJET-V5I3277.pdf>
- [7] <https://www.sciencedirect.com/science/article/abs/pii/S0957417417306577?via%3Dihub>