



## FILM ENDORSEMENT SYSTEM

J. RAJASEKHAR<sup>1</sup>, CH. CHARAN<sup>2</sup>, S. SRIKAR<sup>3</sup>

<sup>1</sup>Assistant Professor, Department of CSE, ST. PETER'S ENGINEERING COLLEGE, HYDERABAD, Telangana, India.

<sup>2,3</sup> UG Scholars, Department of CSE, ST. PETERS ENGINEERING COLLEGE, HYDERABAD, Telangana, India.

**Abstract**— Endorsement means seal of approval. Film endorsement system means to efficiently the ratings for items and recommend items, using the data generated by users. The size of the websites rapidly increasing due to immense growth of internet on these websites it became difficult to find the rating of a film and also recommend according to our choice because it is time consuming process. Recommender systems are used in a variety of areas, with commonly recognized examples taking the form of playlist generators for video and music services, product recommenders for online stores, or content recommenders for social media platforms and open web content recommenders. A recommendation system sometimes replacing system with a synonym such as platform or engine, is a subclass of information filtering system that seeks to predict the “rating” or “preference” a user would give to an item. Collaborative filtering approaches build a model from a user's past behavior as well as similar decision made by other users. Content-based filtering approaches utilize a series of pre-tagged characteristics of an item in order to recommend additional items with a similar property. In this project we have to use a new algorithm that is a FLASK algorithm. For example, companies like Amazon or Netflix use KNN and FLASK when recommending books to buy or movies to watch.

There was even an approximately \$1 million award on Netflix to the team that could come up with the most accurate recommendation algorithm.

**Keywords**— **FLASK, KNN, Content-based filtering**

### I. INTRODUCTION

Man-made consciousness has a sub field called (AI). The point of AI is to grasp the idea

information and fit the information into models that individuals can fathom and utilize. AI is a part of a software engineering that is particular from regular scientific strategies. Calculations are assortments of explicitly coded directions utilized by PCs to gauge or take care of issues in regular processing. AI calculations, then again empower PCs to prepare on information data sources and utilize measurable investigation to deliver values that are in specific reaches. In this project we are using two filtering methods of recommendation system process. We have to check the programming languages which are mostly used for the implementation of the machine learning techniques. By using this programming languages. We get the results either in positive or the negative way for each and every project.

### II. LITERATURE SURVEY

People depend on experience to make decisions that are in their best interests. Recommendation systems are a part of everyday life. Despite the fact that many approaches have been established in the past, search continues to exist due to its widespread use in many applications that personalize recommendations and deal with knowledge overload.

As a result of these demands, various methods such as memory-based and model-based are used. To become a stronger system, the recommendation system also needs to be improved. Amazon.com, Movie-lens, and last .

There are several examples of recommendation systems that provide users with ideas about items that might be of interest to them. Different approaches with their techniques are discussed in this paper in order to

compare the limitations of and technique in order to provide appropriate future recommendations.

### III. EXISTING SYSTEM

There are many recommendation systems which are developed in last few years. We have improved the existing system because in existing system they have used only filtering's with which they have got a less accuracy. In 2018 one group of students have designed the recommendation system using only filtering.

### IV. PROPOSED SYSTEM

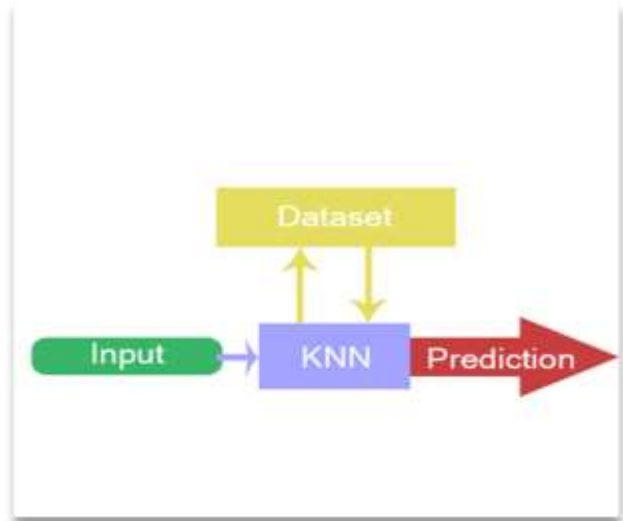
The reason behind this improvement is the popularity gained by organizations like Netflix whose primary objective is customer satisfaction. Before existence the recommendation system, individuals would physically choose movies to watch from movie libraries. They either had to read the user's reviews and based on the review they would select a movie or had to randomly select a movie. This procedure isn't feasible, as there is an enormous number of spectators with a unique preference for movies. Hence many recommendation systems have been developed over the past decade.

In this we are going to recommend movies using 2 types of filtering's they are content-based and collaborative filtering's and KNN algorithm.

### V. ARCHITECTURE DIAGRAM

The tax's diagram represents the flow of the projects such that the steps that are performed. There are two databases such as films data set and the user ratings data that are taken as the source from the internet.

The end user will give the user input to the front-end user interface then by using the flask server the operations are performed to call the back-end server, then after calling the back-end operation content technique is used on the films data and user's data after performing specific operations on the data present, they provide the output at the back end then the output is returned in the front-end application then user can again give another input of the film.



**Dataset:** It is a dataset must be present in every input data file to be shown in prediction page list, otherwise it shown an internal error. Because when we upload such as movie names in the dataset it could be show an output.

The dataset may be used by researchers to validate recommender systems or collaborative filtering algorithms. It may serve as a testbed for matrix and graph algorithms, including PCA and clustering algorithms.

**Input:** A recommendation system provides suggestions to the users through a filtering process that is based on user preference and browsing history. The information about the user is taken an input. The information is taken from the input that is in the form of browsing data.

**Prediction:** Prediction system is an output format for searching as movie name recommendation that shown as a prediction set of related movies search tool. It is main part of this project and useful like everything such a good basics.

Prediction phase it recommends or predicts whatkind of items the user may prefer. This can be made either directly based on the dataset collected in information phase which could be memory based or model based through the system's observed activities of the user.

Recommender systems rely on different types of input such as the most convenient high quality explicit feedback, which includes explicit input by users regarding their interest in item or implicit feedback by inferring user preferences indirectly through observing user behavior. Hybrid feedback can also be obtained through the combination of both explicit and implicit feedback.

Different items to the ones the user chose in the past will be recommended to the user in a content-based approach, while items that similar group people of similar tastes and interests will be recommended in a collective filtering approach. Hybrid systems, which incorporate both approaches in some way, are proposed to address the drawbacks of both approaches.

Shared shifting frameworks assemble client input as evaluations for things in a particular field and use likenesses in rating conduct among numerous clients to decide how to recommend a specific thing. Clients are suggested an item by shared sifting frameworks dependent on the perspectives on different clients.

Memory approach techniques evaluate all user or item data in real time to quantify recommendations and can be divided into three categories: Content-Based (CB) techniques, CF techniques, and hybrid techniques are all available. C F techniques suggest products that have previously been used by similar users; they are based on social, community-driven data.

A theoretical techniques of user rating behavior is proposed in model-based C F algorithms. Instead of using the raw rating data to make predictions, the model's parameters are calculated using the available rating data, and the model is then used to make predictions. Over the last few years, several model-based CF algorithms have been investigated.

One of the most useful machine learning techniques is Naive Bayes. This is due to two factors. For starters, it's simple to implement in software, and it's typically as accurate as more complicated algorithms. Two separate experiments were carried out. The original data was used in the first, which meant that the attribute values were not translated in any way.

Hybrid based filtering is a traditional recommendation method approaches like collaborative filtering ( CF), content-based, and knowledge-based filtering all have their own set of advantages and disadvantages.

## VI.IMPLEMENTATION

Importing pandas, NumPy, matplotlib and sea born are the tools to prepared for the implementation of the project. We have two data sets for the implementation of the project, one is movies data set which contains the details of the movies, another data set is on user ratings for the particular movies this will be used for the collaborative's filtering purpose. Loading the two data sets by using panda's

operations, and retrieving the data set of the movies by using the head () operation.

### NumPy:

This is used to perform the mathematical operation on the file we have imported. NumPy is a two-dimensional exhibit arranged registering library intended for undeniable level numerical capacities and logical calculations. Numpy can be utilized to bring information into the note pad.

### Matplotlib:

Matplotlib is a function used for the data visualization of the project. This is also called as the magic functions which creates the default graphs and changes when we give the data set to the machines.

### Fuzzy wuzzy:

Fuzzy wuzzy is a Python library that is used to fit strings. The method of finding strings that fit a given pattern is known as fuzzy string matching. The differences between sequences are calculated using Liechtenstein Distance.

Seat Geek, a service that helps people find sports and concert tickets, created Fuzzy wuzzy and made it open-source. To problematically match similar data, fuzzy matching is an approximate string- matching technique. Instead of relying on string equis'alence to decide whether two strings are the same, fuzzy matching algorithms attempt to measure how similar two strings are to one another. They may use this information to calculate the probability that two separate strings were intended to be equivalent.

Seat Geek created Fuzzy wuzzy, an open-source string matching library for Python developers, to help determine if two ticket listings of similar names were for the same case. To do so, Fuzzywuzzy uses the Liechtenstein distance (a form of edit distance that takes into account character insertions, deletions, and substitutions).

While fuzzy matching is far from perfect, it can give app developers a place to start when it comes to cleaning up cluttered datasets and creating a more resilient, flexible, and Intuit's user experience. Fine-tuning a fuzzy matching implementation almost always necessitates a lot of thinking and a variety of fuzzy matching techniques. The juice is definitely worth the squeeze for any programmer that must analyses user text input, or for a data set where duplicate entries are an sever-present problem.



## Implementation using K nearest neighbor algorithm:

After implementing the loading process of movies data set and user ratings data set, we have already preferred ruled the content-based filtering, now we have to work on the k nearest neighbor algorithm.

First, we have to check the shape of the both the data sets, the movies data set and the ratings data set. This indicates that number of rows and columns used in this data set.

```
ratings_gf.shape, movies df . shape
```

Now, we have to make the pivot table for the user ratings data set on the behalf of the movie id and columns used are user id, both the user id and movie id is used to pivot table and the values used here are the ratings that user gives to the particular movie id. The ratings depend on the user and the movie all the users may not give ratings to all the movies and some users will also not give ratings to the movies.

### Overview of technologies:

The cosine distance between the item's and the user's vectors can be used to calculate the item's choice for the user, we note that a user's vector would have a positive number for actors who often appear in movies the user enjoys, and a negative number for actors the user dislikes.

To use textual data for predictive modelling, the text must first be filtered to extract specific words, a process known as tokenization. These terms must then be converted to integers or floating-point values in order to be used as inputs in machine learning algorithms. Feature extraction is the term for this procedure.

uses the similarities in the features that have to make the decisions. This method is used to give the recommendations on the basis of the content that user gives for example, let us consider film endorsement system this is basically works on the recommendations we can also use this content-based filtering for this example we can get the recommendations of the movie on the basis of the genres that the user gives as the user input to the machine.

If the user gives the horror movie as the input to the machine then it will sort the names of the related movie on the basis of the genre that the user gives.

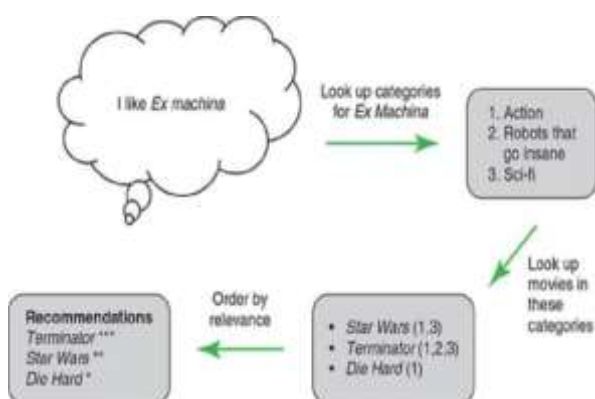
We must create a profile for each object that represents the item's most significant characteristics. If we make a movie as an object, for example, the actors, producer, release year, and genre are the most important aspects of the film. In the Item Profile, we can also provide its IMDB (Internet Movie Database) ranking.

### Result:

Now by comparing all the algorithm content-based filtering is the best algorithm so we have done the front-end application implementation using flask using this algorithm.

By comparing all the outputs, the best output will be the content-based filtering because that recommendation is done by using the genres and the other best output will be the item-based filtering because the movies which are recommended is based on the user's ratings and the movies so by using these algorithms, we have recommended the movies.

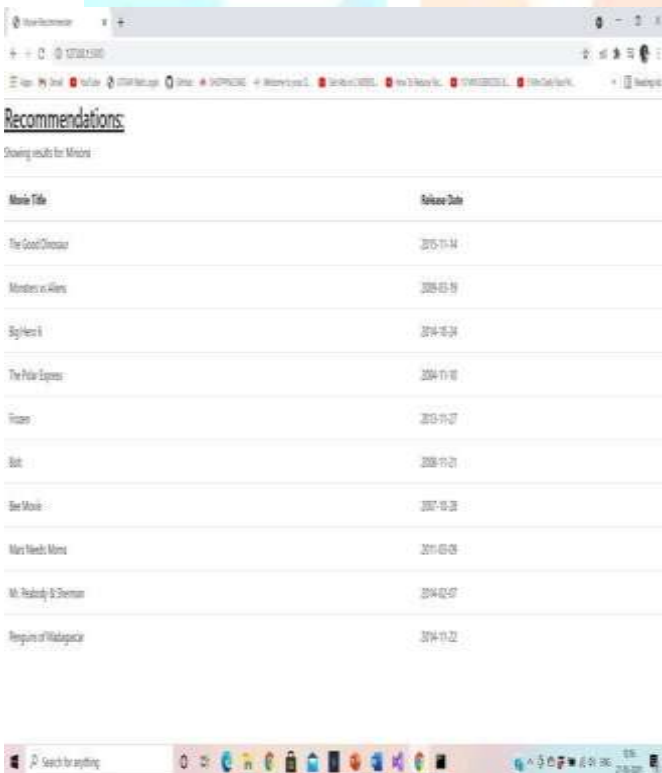
So, here in k nearest neighbors all the data set is used efficiently and gives the good result.



A content-based recommendation system in attempts to guess a user's features or actions based on the features of an object to which he or she responds positively. This



The outputs will be hashing the recommendations of the movies with the release dates of the movies that are recommended. Here in the above pictures the input is given as avatar and content-based filtering is running in the back-end part and the output is taken to the end by using the flask technology.



## VIII.CONCLUSION

Film endorsement system can be a challenging task due to the high number of attributes that should be considered for the accurate recommendations. The major step in the recommendation process is collection and pre-processing of the data.

This project evaluates the Film endorsement system using the Kaggle data set which gives the recommendations of movies using different machine learning algorithms. The most relevant features used in this recommendation system are genres, id, original title, release date, title, director, cast by filtering out outliers and the irrelevant features of the data set.

Being sophisticated model, Content based filtering and k nearest neighbor's algorithm will give the good and accurate results in comparison to the prior work using these data sets.

Although the system has achieved astonishing performance in Film endorsement system problem our aim for the future research is to test this system to work successfully with various data sets. We will extend our data set with the real time examples such as IMDB and validate the proposed approach.

## REFERENCES

1. Concept of count vectorizer: [https://scikitlearn.org/stable/modules/generated/sklearn.feature\\_extraction.text.count\\_vectorizer.html](https://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.text.count_vectorizer.html)
2. movies data set in retrieved from Kaggle website <https://www.kaggle.com/tmdb/tmdb-movie-metadata>
3. Rating date set is retrieved from Kaggle website: <https://grouplens.org/datasets/movielens>
4. Sea born concepts: <https://seaborn.pydata.org/>
5. k nearest neighbor algorithm concept: <https://scikit-learn.org/stable/modules/neighbors.html>
6. Sparse matrix is concept is retrieved from <https://docs.spicy.org/doc/spicy/references/sparse.html>
7. Fuzzy wuzzy concept: <https://pypi.org/project/puzzywuzzy/>