OPTIMIZATION OF RECOMMENDER SYSTEM WITH NEURAL COLLABORATIVE FILTERING AND GENETIC ALGORITHM

Minor Project II

Submitted by

Divyansh Dhingra (19103077)
Chetanya Kumar Jha (19103081)
Harsh Chauhan (19103085)

Under the supervision of

Ms. Deepti Singh

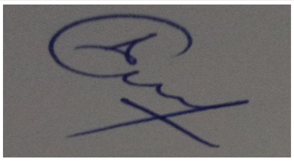
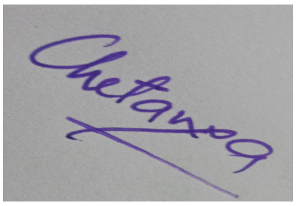
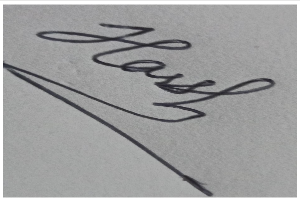
विद्या तत्व ज्योतिसमः

Department of CSE/IT
Jaypee Institute of Information Technology University, Noida

MAY 2022

ACKNOWLEDGEMENT

It gives us immense pleasure to express our deepest sense of gratitude and sincere thanks to our highest respected and esteemed guide ***Ms. Deepti Singh*** for her valuable guidance, encouragement and help for completing this work. Her useful suggestions for this whole work and co-operative behavior are sincerely acknowledged.
We would like to express our sincere thanks to her for giving us this opportunity to undertake this project. We would also like to express our indebtedness to our parents as well as our family members whose blessings and support always helped us to face the challenges ahead. We wish to express our sincere thanks to our other classmates for their insightful comments and constructive suggestion to improve the quality of this project work.

Student ID	Student Name	Student signature
19103077	**Divyansh Dhingra**	
19103081	**Chetanya Kumar Jha**	
19103085	**Harsh Chauhan**	

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and beliefs, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma from a university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place: Jaypee Institute of Information Technology, Noida

Date: 20th May 2022

Name: Divyansh Dhingra

Enrolment No.: 19103077

Name: Chetanya Kumar Jha

Enrolment No.: 19103081

Name: Harsh Chauhan

Enrolment No.: 19103085

CERTIFICATE

This is to certify that the work titled "**Optimization of Recommender System using Neural Collaborative Filtering and Genetic Algorithm**" submitted by (Divyansh Dhingra, Chetanya Kumar Jha, Harsh Chauhan) of B.Tech of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Digital Signature of Supervisor Ms. Deepti Singh
Name of Supervisor Ms. Deepti Singh
Designation Assistant Professor
Date 20th May 2022

Table of Contents

Acknowledgement *ii*

Declaration *iii*

Certificate *iv*

CONTENTS

List of Figures *6*

List of Tables *6*

Abbreviations *7*

1. Introduction *8-12*

1.1 Recommendation System and their consequences in the business world *9*

1.2 How does Recommendation System of Amazon/ Prime work? *10*

1.3 Recommendation System Techniques *11-12*

2. Literature Survey *13-14*

3. Project Flow Diagram

15

4. Requirement Analysis *16-17*

4.1 Data Requirements *16*

4.2 Functional Requirements *17*

5. Detailed Design *17-21*

5.1 Neural Collaborative Filtering

18-19

5.2 Genetic Algorithm - Selection, Crossover, Mutation *20*

5.3 Activation Functions Used *21*

6. Implementation *22-23*

7. Experimental Results & Analysis *23-26*

8. Conclusion & Future Scope *27*

9. References *28-29*

List of Tables

Table	Title	Page
1	Some popular sites that use Recommendation Systems………...................	09
2	Different Types of Recommendation Techniques…………..	11
3	Comprehensive Analysis of literature surveys...	13-14
4	Comparative & Evaluative analysis of different algorithms implemented	26

List of Figures

Figure	Title	Page
1	Growing Recommendation System Market Size over years	10
2	Flow Representation of various Recommendation System Technique.............	12
3	Flow Representation of our proposed algorithm..	15
4	Histogram of ratings by users..	17
5	Neural Collaborative Filtering Flow Diagram...	18
6	Artificial Neural Network structural representation..	19
7	Genetic Algorithm structural representation………………………….........	20

Abbreviations

NCF	Neural Collaborative Filtering
RS	Recommendation System
GA	Genetic Algorithm
ANN	Artificial Neural Network
RMSE	Root Mean Square Error
MSE	Mean Square Error
CF	Collaborative Filtering
CB	Content-Based
SVD	Singular Value Decomposition

1. Introduction

The growth and distribution of the Internet and smart devices have resulted in a considerable rise in traffic to online, app, and social media platforms. Furthermore, these platforms are expanding their gathering of diverse data that may be used to determine users' preferences. The active use of users' platforms, in particular, allows for the collection of a wide range of data, including information about the user's followers, tweet data, and material supplied by the user. As a result, the Internet and smart gadgets have evolved into ecosystems where diverse user data may be gathered.
Recommender Systems are software tools and approaches that provide recommendations for items that a user would find useful. The recommendations are intended to assist users in different decision-making processes, such as deciding what to buy, what music to listen to, or what news to read. Recommender systems have shown to be an effective way for internet users to cope with information overload, and they have grown in popularity to become one of the most powerful and widely used tools in electronic commerce. As a result, several strategies for suggestion generation have been presented, and many of these have been effectively used in business situations during the previous decade. Individuals who lack adequate personal experience or skill to assess the potentially overwhelming amount of alternative things that a Web site, for example, may provide are the primary target of RS's. The creation of RS began with a simple observation: people frequently depend on suggestions from others while making normal, daily decisions.

As evidenced by the following data, interest in recommender systems has skyrocketed in recent years:

- Such highly rated Internet sites as Amazon.com, YouTube, Netflix, Yahoo, and IMDB use recommender systems extensively. Furthermore, several media businesses are also creating and deploying RSs as part of their subscriber offerings. For example, Netflix, the online movie rental business, presented a million-dollar reward to the team that initially improved the effectiveness of their recommender system by a significant margin.
- In the fields of databases, information systems, and adaptive systems, RS sessions are regularly incorporated into more traditional conferences. ACM SIGIR Special Interest Group on Information Retrieval (SIGIR), User Modeling, Adaptation, and Personalization (UMAP), and ACM's Special Interest Group on Data Management are among these conferences (SIGMOD).

There are myriad reasons why service providers may want to exploit & explore this technology:

- ***Increase the number of items sold:*** This is perhaps the most crucial capability for a commercial RS, namely, the ability to offer a different group of things than those that are normally sold without any sort of suggestion. This objective is met since the suggested goods are likely to meet the user's needs and desires.
- ***Sell more diverse items:*** Another important feature of an RS is that it allows the user to choose goods that would otherwise be difficult to locate without a specific suggestion.

- ***Increase user satisfaction:*** A well-designed RS may also enhance the user's experience with the website or application. The user will find the recommendations fascinating and relevant, and she will like using the system if the human-computer interface is well-designed. The combination of effective recommendations with a user-friendly interface will improve the system's subjective rating. As a result, system utilization will grow, as will the probability that the recommendations will be accepted.
- ***Increase user fidelity:*** A user should be devoted to a website that identifies and honors returning customers as important visitors when they arrive. This is a common element of an RS since many RSs compute suggestions based on data gathered from the user in past encounters, such as her item ratings.

Table 1. Some Popular Sites that Use Recommendation Systems

Sites	Recommendations
Netflix	Movies, TV Series, Web Series
YouTube	Videos, Shorts,
Amazon	Consumer Products
Facebook/ Instagram	Friend Suggestions, Videos, and Posts
CareerBuilder	Jobs

1.1 Recommendation Systems & their consequences in the business world

Companies like Amazon and Netflix, are recognized for their individualized consumer experiences and use of recommendation algorithms as crucial and beneficial tools. Customers' demographic data is collected and analyzed by each of these firms, which is combined with information from prior purchases, product evaluations, and user activity. eCommerce companies may locate opportune periods to offer new things that you're likely to buy by employing various "filtering" methods.

The worldwide recommendation engine market was worth USD 1.77 billion in 2020, and it is predicted to increase at a CAGR of 33.0% from 2021 to 2028. The need for recommendation engines is growing in response to the growing desire to improve customer experience. The increased need for recommendation engine solutions is a result of enterprises' increasing embrace of digital technology. These effects are projected to linger much beyond the epidemic, affecting both companies and individuals. For example, in the first quarter of 2020, the e-commerce behemoth Amazon.com, Inc. earned about USD 33 million in sales every hour.

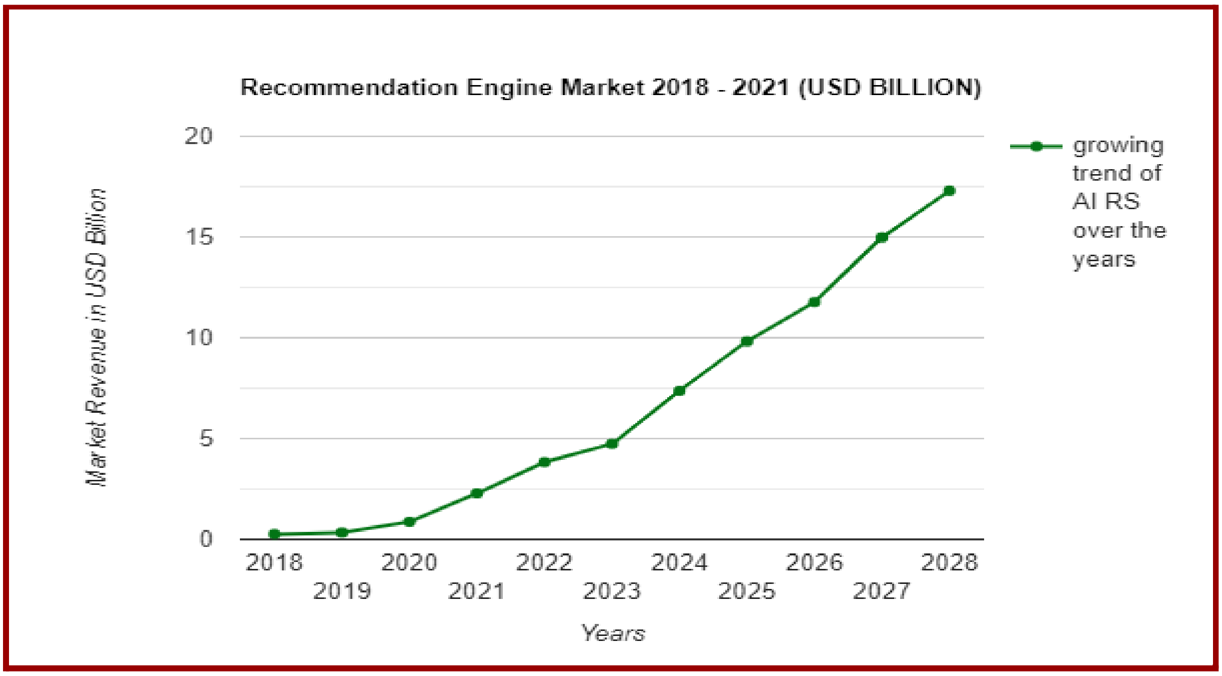


Figure 1. Growing Recommendation system market size over the years

1.2 How does RS of Amazon/ Prime work?

Most customers are unable to subscribe to all of the direct-to-consumer channels accessible today due to the ever-increasing number of options? Content and user experience both influence subscription/purchase decisions. When it comes to deciding what to buy, how to buy it, and how to connect with it, today's consumers want real-time, customized experiences. Whether it's raising click-through rate, increasing views, view length, subscriptions, or premium content sales, media firms are always looking for new methods to improve the customer experience and increase profits. Recommender systems are an important tool for achieving these objectives. DTC platforms can keep users engaged after they've seen the material that first drew them to the platform by providing suggestions that optimize the value of extensive content catalogs.

Amazon.com Recommendations: Item-to-Item Collaborative Filtering" - The most frequent method of internet product suggestion is collaborative filtering. It's "collaborative" because it anticipates a customer's choices based on the preferences of others. The world was centered on collaborative filtering based on user input. When a person visits the website, they want to recommendations her users are similar to them. We flipped it on its head and came up with a new method that had far better scalability and quality features for online recommendations." The superior approach was to base product suggestions on product correlations rather than client similarities. A visitor to Amazon.com would be paired with other customers who had comparable purchase histories, and those purchase histories would propose recommendations for the visitor using user-based collaborative filtering. Studying purchase histories at the item level provided better suggestions than analyzing them at the customer level, according to Amazon's Personalization team. To create a new Prime Video recommendation system. Matrix factorization, which discovers relatively tiny matrices that, when multiplied together, mimic a much bigger matrix, was the conventional approach for producing individualized suggestions at the time.

1.3 Recommendation System Techniques

Recommendation System must anticipate that an item is worth suggesting to carry out its fundamental role of discovering valuable items for the user. To do so, the system must be able to anticipate the usefulness of some of them, or at the very least compare the utility of some of them, before deciding which products to recommend based on the comparison. Table 2 shows the anatomy of different recommendation filtering techniques.

Table 2. Different types of recommendation system techniques

Type	Explanation	Example
Content-Based	The algorithm learns to suggest goods that are similar to those that the user has previously enjoyed. The attributes associated with the compared items are used to compute item similarity.	News Dude is a personal news system that utilizes synthesized speech to read news stories to users. LIBRA is a content-based book recommendation system
Collaborative Filtering	The most basic and original version of this strategy suggests products to the active user that other users with similar likes have previously enjoyed. The closeness in taste between two users is determined by the individuals' rating histories being comparable.	Amazon.com is an example of an e-commerce recommendation engine that recommends online items for various consumers using scalable item-to-item collaborative filtering algorithms.
Knowledge-Based	Knowledge-based systems make recommendations based on domain knowledge about how various item attributes match users' wants and preferences, and, ultimately, how the item is helpful to the user.	Atlassian Confluence is a cloud-based application that delivers a full customer self-service and case support suite when combined with Atlassian's Jira Service Management product.
Demographic	This sort of technology suggests goods depending on the user's demographic profile. Distinct suggestions should be provided for different demographic niches, according to the ademption.	Tofulfillmarketers efficiently target their customers, Facebook and Instagram utilize highly detailed demographic segmentation.
Community-Based	This type of technology suggests things depending on the user's friends' tastes. This method is based on the epigram "Tell me who your friends are, and I'll tell you who you are."	To simulate the generic network of interests, the Internet Movie Database (IMDb) uses the underlying social network graph of the movies based on their shared reviews.
Hybrid Recommender Systems	These RSs are built using a mixture of the procedures discussed above. A hybrid system that combines approaches A and B aims to leverage the benefits of A to compensate for the shortcomings of B.	The PTV system creates a user's TV viewing schedule by integrating suggestions from content-based and collaborative systems.

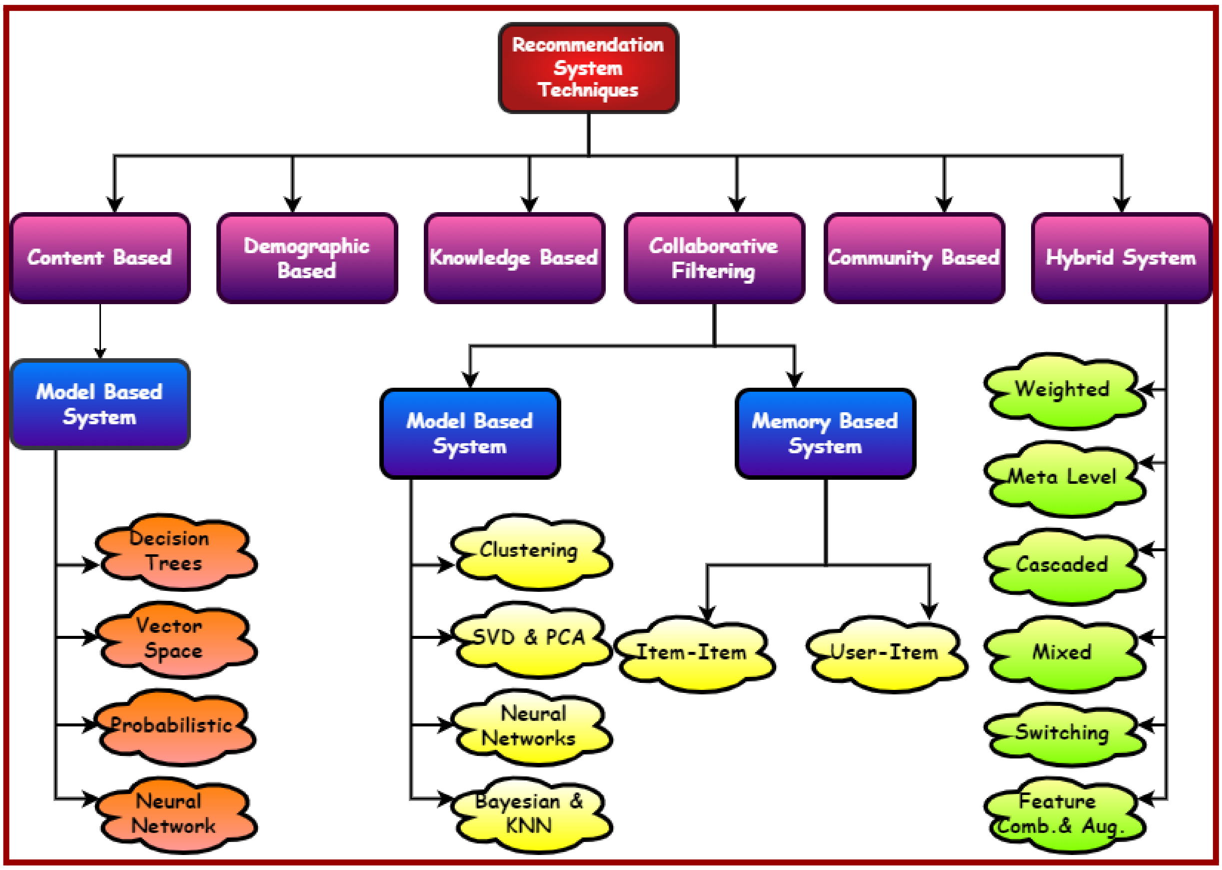


Figure 2. Flow Representation of various Recommendation Techniques

2. Literature Survey

Recommender systems are capable of providing users with personalized and personalized recommendations. As a result, it addresses the problem of information overload that consumers face in today's world. Various ways to develop recommendation systems have recently been proposed. Despite their broad popularity, the two filtering approaches have several drawbacks. While Content-Based Filtering techniques have limitations such as regional content analysis, overspecialization, and data sparsity, collaborative approaches have cold-start, sparsity, and scalability concerns. These difficulties make using these systems in live production challenging. To address these challenges, hybrid filtering has been proposed, which combines two or more filtering approaches in different ways to improve the performance and accuracy of recommender systems. These hybrid systems attempt to combine the advantages of each technique without compromising any functionality owing to inherent flaws.

The Literature survey comprises various researches done or is ongoing in the particular domain. Various approaches, algorithms, and solutions have been provided and have tried to fulfill the gaps created be it in terms of accuracy, errors, losses, data sparsity, cold start, data over specialization, data overfitting, and underfitting, etc. The solution proposed by us tries to fill the gaps yet requires further changes and commutations.

The detailed overview of various survey papers, IEEE transactions, Tutorials, etc has been delineated henceforth in a tabulated form:

Table 3. Literature Review of different research papers

S. No.	Author	Year	Contributions	Algorithm	Datasets Used/ Applications
1	Oumaima Stitini, Soulaimane Kaloun, Omar Bencharef	2022	An Improved Recommender System Solution to Mitigate the Over-Specialization Problem using Genetic Algorithm	Genetic Algorithm, Content-Based Filtering	Movielens-10M dataset, tackled overspecialization problem in content-based RS
2	Abbas Ali Rezaee, Navid Abravan	2020	A hybrid friend-based recommendation system using the combination of Meta-heuristic Invasive weed and genetic algorithms	Invasive Weed Algorithm, Genetic Algorithm	It proposes Euclidean, Minkowski and Jaccard methods better classes user's similarity. Movielens Dataset is used
3	Md Rafidul Islam Sarker, Abdul Matin	2021	A hybrid Collaborative Recommendation System Based on Matrix Factorization And Deep Neural Network	Deep Neural Networks, Mtrix Factorization, Similarity Based Collaborative Filtering, NHF	Movilens Dataset, Netflix Dataset, Sigmoid AF is used
4	Guoshuai Wei, Quanwang Wu, MengChu Zhou	2021	A Hybrid Probabilistic Multiobjective Evolutionary Algorithm for Commercial Recommendation Systems	MOEA-ProbS, PMOEA, MOEA-EPG, HP-MOEA	Movielens Dataset, Netflix Dataset, Book-Crossing Dataset
5	Rahul Katarya, Om Prakash Verma	2016	Recommender system with grey wolf optimizer and FCM	Collaborative filtering, Gray wolf optimizer, Fuzzy c-mean	ML-1M, ML-10M
6	Michael Farber, Adam Jatowt	2020	Citation Recommendation: approaches and datasets	Neural Network (LSTM), Neural Network (Feed Forward)	CiteSeerX, RefSeer
7	Yehuda Koren, Robert Bell	2019	Advances in Collaborative Filtering	Matrix Factorization, SVD, SVD++	It delineates the detailed description of methods to tackle challenges of Netflix Prize Competition
8	Jasminder Kaur Sandhu, Deepam Goyal, Anjali	2021	User Profiling in Travel Recommender System using Hybridization and Collaborative Method	Fuzzy C Clustering, Location Based Services and Networks	Purpose of TRS is to search the information of traveler and gain confidence in providing recommendations

9	Ashkan Yeganeh Zaremarjal	2021	Semantic Collaborative Filtering Recommender System Using CNNs	Convolutionsl Neural Network, Auto Encoders, LSTM	Jester Dataset3, Jester Dataset4
10	Siddhi Khanse, Payal Bhandri, Atharva Dharane	2020	Comparitive Study of Genetic Algorithm and Artificial Neural Network for Multi-class Classification based on Type-2 Diabetes Treatment Recommendation model	Genetic Algorithm, Multi-class Classification ANN	American Diabetes Association, Type-2 Diabetes Patient Data
11	Maram Almaghrabi, Girija Chetty	2020	Multilingual Sentiment Recommendation System based on Multilayer Convolutional Neural Networks (MCNN) and Collaborative Filtering based Multistage Deep Neural Network Models (CFMDNN)	NLP, MCNN, CFDMNN	Booking Hotel Dataset, Arabic Hotel Review Dataset
12	Yimin Peng, Rong Hu, Yiping Wen	2021	CA-NCF: A Category Assisted Neural Collaborative Filtering Approach for Personalized Recommendation	Neu-MF, Optimized NCF, NDCG	Provides optimized Neural Collaborative Filtering Framework for item recommendation
13	Neelika Chakrabarti, Sheona Das	2019	Neural Netowrks and Collaborative Filtering	Collaborative Deep Learning, Recurrent Neural Networks, Collaborative Topic Regression	Movielens Dataset, Netflix Dataset
14	Jing Yu, Jingjing Shi, Yunwen Chen	2021	Collaborative Filtering Recommendation with Fluctuations of User' Preference	KNN, UCF, OCF, ACOS, CPCC	MovieLens 1M and MovieLens latest small
15	Bin Li, Hua Xia, Sailuo Wan	2020	The Research for Recommendation System Based on Improved KNN Algorithm	IKNN, TF-IDF,	ML-100K, ML-1M, ML-10M

3. Project Flow Diagram

In this section, we are proposing the Project-Flow diagram of our working model.

Figure 3. Flow Representation of our proposed algorithm

4. Requirement Analysis

The process of developing user expectations for new or upgraded software is known as requirement analysis. The tasks that go into determining the needs or conditions to meet for a new or altered product or project, taking into account the potentially conflicting requirements of various stakeholders, analyzing, documenting, validating, and managing software or system requirements are all included in the requirements analysis. Functional needs, non-functional requirements, and typical, anticipated, and sensational requirements are the distinct sorts of requirements. To build this project and there are certain requirements that needed to be fulfilled for the successful implementation. The requirements are as follows:

4.1 Data Requirements

The process of identifying, prioritizing, precisely formulating, and validating the data required to meet the objectives are defined as data
requirements. For this project, the dataset which we have worked on is GroupLens - Movielens Dataset - ML-10M which comprises

4.1.1 Movies.csv - MovieId, Title, Genre

```
movie_data.head()
```

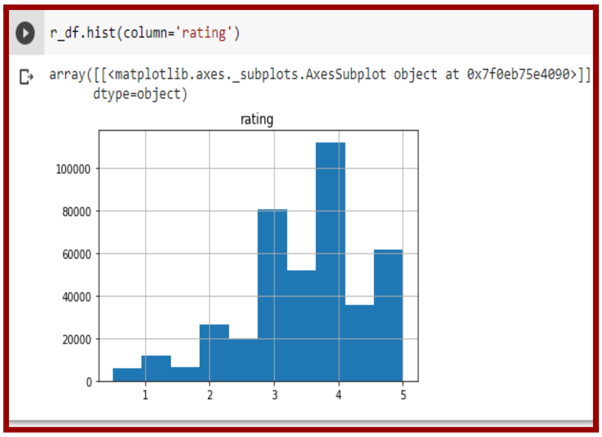
	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

4.1.2 Ratings.csv - User Id, MovieId, Ratings, Timestamp

```
rating_data.head()
```

	userId	movieId	rating	timestamp
0	1	296.0	5.0	1.147880e+09
1	1	306.0	3.5	1.147869e+09
2	1	307.0	5.0	1.147869e+09
3	1	665.0	5.0	1.147879e+09
4	1	899.0	3.5	1.147869e+09

4.1.3 Histogram of Ratings by users depicting the most number of ratings by user



4.2 Functional Requirements

Functional Requirements explain what the system will accomplish, generally in terms of functions it should be able to do.

4.2.1 Tools, Technologies & Libraries

- **Visual Studio Code**:- Visual Studio Code is a code editor redefined and optimized for building and debugging modern web and cloud applications.
- **Tensorflow** - TensorFlow provides a collection of workflows to develop and train models using Python.
- **Keras**:- Keras is a powerful and easy-to-use free open-source Python library for developing and evaluating deep learning models.
- **NumPy**:- NumPy can be used to perform a wide variety of mathematical operations on arrays.
- **Pandas**:- Pandas are mainly used for data analysis and associated manipulation of tabular data in Dataframes.
- **Sklearn**:- Python Library used for machine learning and statistical modeling including classification.
- **Matplotlib**:- Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- **Seaborn**:- Seaborn is a Python data visualization library based on matplotlib.
- **Datetime**:- Datetime module supplies classes for manipulating dates and times.
- **Math**:- Math Library provides us access to some common math functions and constants in Python.

URL:- https://grouplens.org/datasets/movielens/25m/

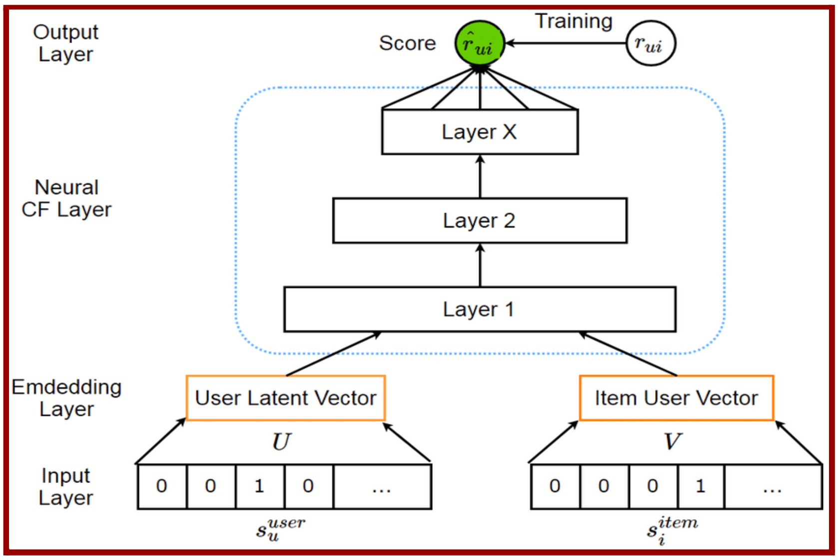
5. Detailed Design

The following are some of the stages involved in this project:

- The dataset is separated into training and testing data after pre-processing. 75 percent of the rows in the balanced dataset will be utilized to train the ANN model, while the remaining 25% will be used for testing.
- An artificial neural network is used to anticipate a user's rating of a particular film.
- In order to enhance movie rating prediction, a Genetic Algorithm is used to develop an optimum topology for Artificial Neural Network.
- The accuracy of the constructed model is then determined by comparing predicted and real movie ratings.

5.1 Neural Collaborative Filtering

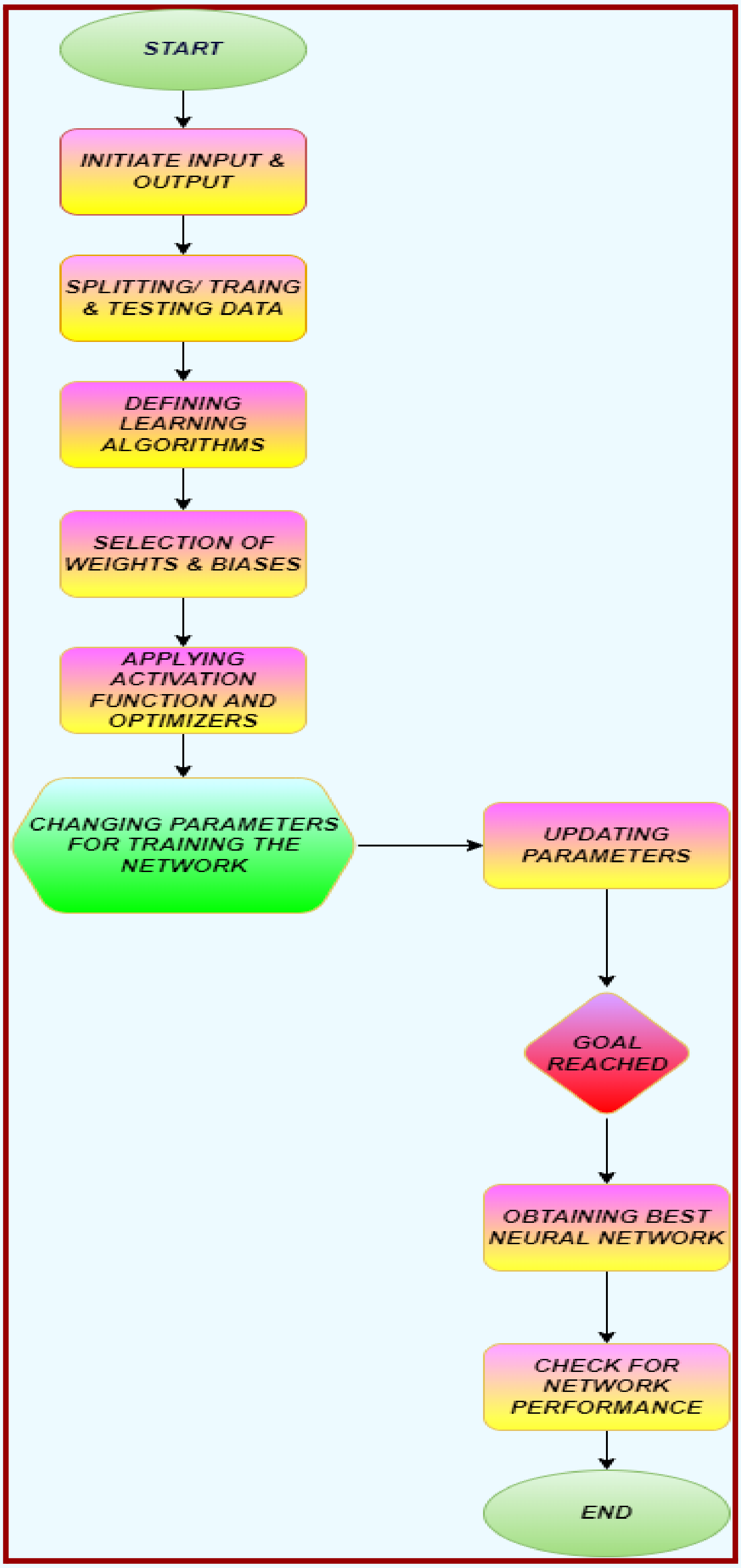
NCF is a generic framework that can express and generalize matrix factorization. We suggest using a multi-layer perceptron to train the user-item interaction function to boost NCF modeling with non-linearities. Deeper layers of neural networks provide greater recommendation performance, according to empirical data.



(Source: Zhang, Shuai & Yao, Lina & Sun, Aixin & Tay, Yi. (2017). Deep Learning-Based Recommender System: A Survey and New Perspectives. 10.1145/3285029.)

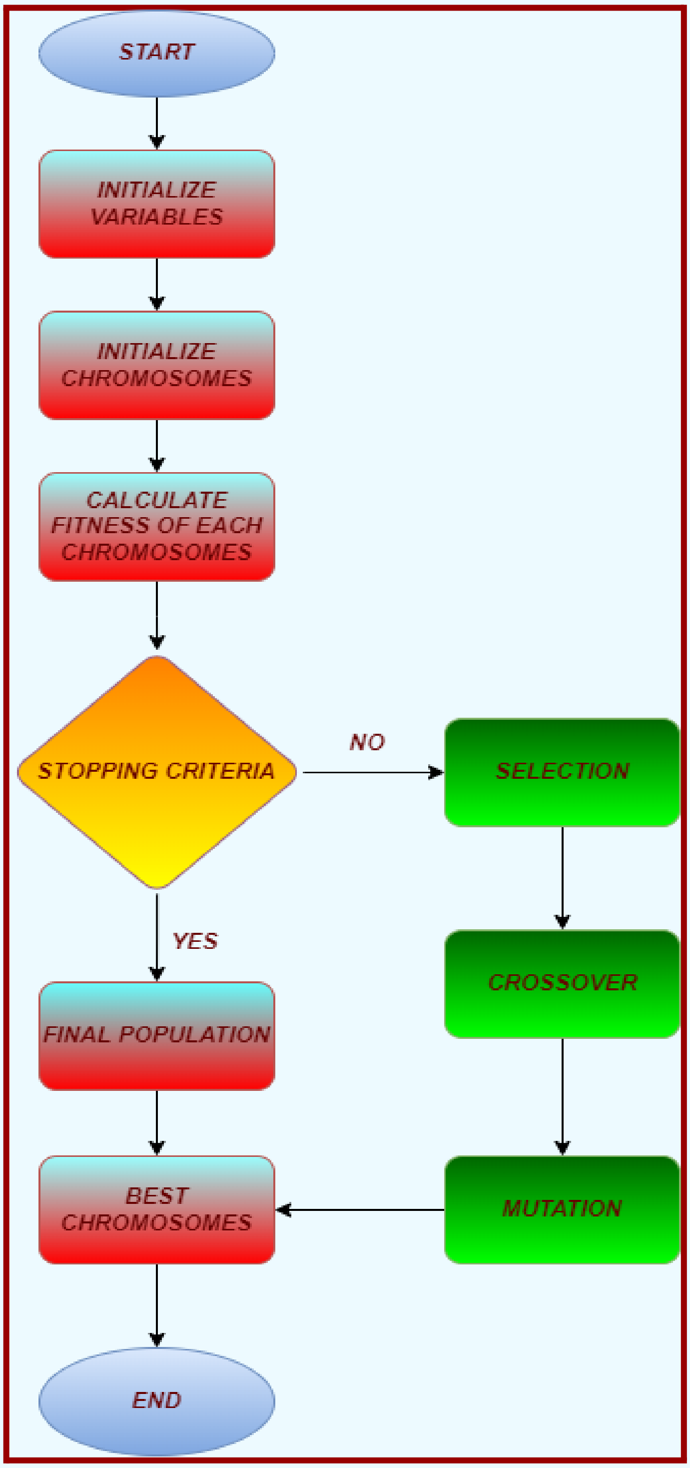
- Binarise a sparse vector for the user and object identification in the Input Layer, where: Item I: 1 indicates that the user u has interacted with Item i to locate the user
- The embedding layer is a completely linked layer that transforms a sparse representation into a dense vector. The latent user/item vectors are the user/item embeddings obtained.
- Multi-layered neural architecture is used by neural CF layers to map latent vectors to prediction scores.
- By reducing the pointwise loss/pairwise loss, the final output layer delivers the anticipated score.

Artificial Neural Networks (ANN) are multi-layer fully connected neural nets that look like this diagram. These layers consist of an input layer, many hidden layers, and an output layer. Each node in one layer is connected to those in the next. We may make the network deeper by increasing the number of hidden layers.



5.2 Genetic Algorithm

Genetic algorithms are a type of search technique that is used in computers to discover an exact or approximate answer to optimization and search issues. Global search heuristics are another title for them. One of the first population-based stochastic algorithms proposed in history is the Genetic Algorithm. Inheritance mutation, selection, and cross-over are all approaches inspired by evolutionary biology. GA's major operators, like those of other EAs, are selection, crossover, and mutation. These techniques provide a method for programs to enhance their settings automatically. The new populations are created by repeatedly applying genetic operators to existing individuals in the population.



5.3 Activation Function Used:

5.3.1 ReLU Activation Function

The Rectified Linear Unit is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input, but for any positive value x, it returns that value back. So it can be written as f(x)=max(0,x).

5.3.2 Tanh or hyperbolic tangent Activation Function

A displaced sigmoid neuron is what it is. It essentially squashes a real-valued number between -1 and +1. It saturates at large positive and negative values, just like the sigmoid neuron. Its output is always zero-centered.

The function is: $$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Its derivative is: $$\frac{d}{dx}\tanh(x) = 1 - \tanh^2(x)$$

5.3.3 Softmax Activation Function

Softmax is a vector function at its core. It accepts a vector as input and outputs another vector; in other words, it has many inputs and outputs.

5.3.4 Sigmoidal Activation Function

It takes a real-valued number as input and compresses it between 0 and 1. It is also referred to as a squashing function. Its goal is to add nonlinearity to the input space. We receive the wiggle and the network learns to record intricate relationships because of the non-linearity.

The function is: $$S(x) = \frac{1}{1+e^{-x}}$$

Its derivative is: $$\frac{d}{dx}S(x) = S(x)(1 - S(x))$$

6. Implementation

6.1 Dataset Preprocessing

	movieId	title	genres	year	Adventure	Animation	Children	Comedy	Fantasy	Romance	...	Horror	Mystery	Sci-Fi	IMAX	Documentary	War	Musical	Western	Film-Noir	(no genres listed)
0	1	Toy Story	[Adventure, Animation, Children, Comedy, Fantasy]	1	1.0	1.0	1.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	2	Jumanji	[Adventure, Children, Fantasy]	1	1.0	0.0	1.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3	Grumpier Old Men	[Comedy, Romance]	1	0.0	0.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	4	Waiting to Exhale	[Comedy, Drama, Romance]	1	0.0	0.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	5	Father of the Bride Part II	[Comedy]	1	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Populations are generated randomly

```
generate_chromosome(15)
[23] ✓ 0.1s
... '000100110000000'

generate_population(15,5)
[24] ✓ 0.1s
... ['00110',
 '01001',
 '10011',
 '01101',
 '10111',
 '01101',
 '11011',
 '00011',
 '00101',
 '00110',
 '00010',
 '01001',
 '01001',
 '11000',
 '01000']
```

After selection, single-point crossover, binary encoding, and mutation the population which remains is

```
s = generate_population(4,10)
print(s)
p = mut(s)
print('\n',p)
✓ 0.1s
['1011101001', '1111100000', '1100000001', '1100010000']

 ['1001101001', '1111100000', '1100000001', '1100010100']
```

The best chromosomes selected using roulette wheel selection with a number of genes in chromosomes = 4 and the total population 10.

```
p=generate_population(10,4)
print(p)
f=calculate_fitness(p)
p2=roulette_wheel_selection(p,f)
print(p2)
✓ 0.1s
['1010', '1110', '1111', '1010', '1111', '0001', '1101', '1100', '0111', '0001']
['1100' '1111' '1110' '0001' '0001' '1010' '1010' '0001' '1110' '1111']
```

A number of neurons and activation functions used in fully connected hidden layers are found using the genetic algorithm. Optimized topology of Artificial Neural Network:

```
optimum_ann_topology(population_size, chromosome_length, generations, mutation_prob)
✓ 0.1s
'0101011110110101101111011'
```

A multitude of factors and hyperparameters impact the performance of a neural network. These factors influence the majority of ANN output. They include weights, biases, learning rate, batch size, and other parameters. The weight of each node in the ANN is different. Each node in the network is given a weight. A transfer function is used to compute the weighted total of the inputs as well as the bias. After the transfer function has computed the sum, the activation function obtains the outcome. Based on the output received, the activation functions fire the appropriate result from the node. Artificial Neural Networks use sigmoid, RELU, Softmax, tanh, and other popular activation functions.

7. Experimental Results & Analysis

Neural Collaborative Filtering (NCF):

• NCF is used to predict the rating given by a user to a particular movie.

• Input layer of NCF contains 6 neurons and the activation function used in the Input layer is tanh.

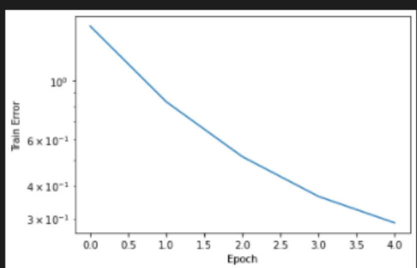
• There are 20, 30, and 50 embedding layerS.

• Output layer of NCF contains 10 neurons. ReLU activation function is applied in the output layer of NCF which will output five probabilities corresponding to ratings from 1 to 5. The rating corresponding to maximum probability will be considered as the final predicted rating.

• MSE is used as the loss function during backpropagation. The learning rate was kept at 0.25 and The of Epochs is 10.

7.1 Using tanh activation function the loss calculated at each epoch is (with embedding size = 20 and epochs = 5)

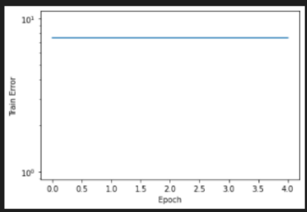
```
Epoch 1/5
2292/2292 [==============================] - 60s 8ms/step - loss: 2.4802
Epoch 2/5
2292/2292 [==============================] - 19s 8ms/step - loss: 0.9069
Epoch 3/5
2292/2292 [==============================] - 19s 8ms/step - loss: 0.5172
Epoch 4/5
2292/2292 [==============================] - 19s 8ms/step - loss: 0.3489
Epoch 5/5
2292/2292 [==============================] - 19s 8ms/step - loss: 0.2657
```



7.2 Using Softmax AF with embedding size 50 and epochs 5

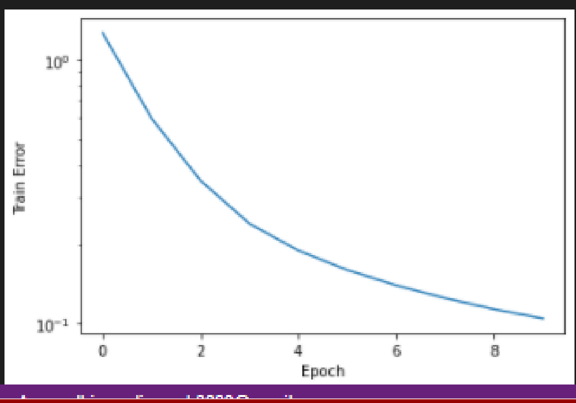
```
model = Neural_CF_Model(embed_size=20)
history = model.fit([train_data['user_encoded'], train_data['movie_encoded']], train_data['rating'], epochs=5)
pd.Series(history.history['loss']).plot(logy=True)
plt.xlabel("Epoch")
plt.ylabel("Train Error")
plt.show()
```

```
Epoch 1/5
2292/2292 [==============================] - 56s 5ms/step - loss: 7.5362
Epoch 2/5
2292/2292 [==============================] - 12s 5ms/step - loss: 7.5232
Epoch 3/5
2292/2292 [==============================] - 12s 5ms/step - loss: 7.5381
Epoch 4/5
2292/2292 [==============================] - 12s 5ms/step - loss: 7.5359
Epoch 5/5
2292/2292 [==============================] - 12s 5ms/step - loss: 7.5120
```



7.3 Using ReLU AF with embedding size 50 and no of epochs = 10.

```
Epoch 1/10
4790/4790 [==============================] - 207s 32ms/step - loss: 2.1119
Epoch 2/10
4790/4790 [==============================] - 149s 31ms/step - loss: 0.6295
Epoch 3/10
4790/4790 [==============================] - 154s 32ms/step - loss: 0.3278
Epoch 4/10
4790/4790 [==============================] - 159s 33ms/step - loss: 0.21271s - loss: 0.21 - ET
Epoch 5/10
4790/4790 [==============================] - 158s 33ms/step - loss: 0.1663
Epoch 6/10
4790/4790 [==============================] - 156s 33ms/step - loss: 0.1418
Epoch 7/10
4790/4790 [==============================] - 158s 33ms/step - loss: 0.1224
Epoch 8/10
4790/4790 [==============================] - 159s 33ms/step - loss: 0.1115
Epoch 9/10
4790/4790 [==============================] - 162s 34ms/step - loss: 0.1011
Epoch 10/10
4790/4790 [==============================] - 156s 33ms/step - loss: 0.09270s - loss:
```



7.4 Metric used to measure the correctness of the recommendation system is RMSE - Root Mean Squared Error -

The root-mean-square error (RMSE) is a commonly used metric for comparing predicted and observed values by a model or estimator. The RMSE is a measure of how evenly distributed the residuals are. In other words, it indicates how tightly the data is clustered around the line of best fit.

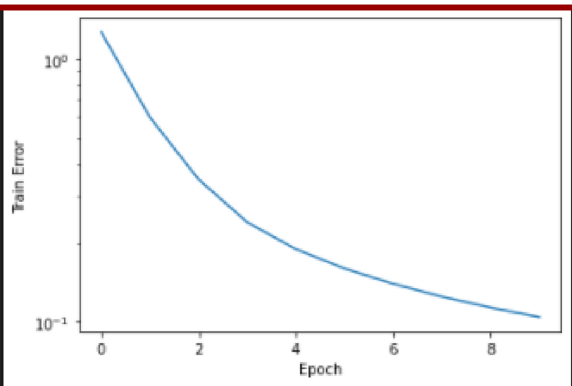
```
d = {'prediction': y_pred.tolist(), 'true_value': y_actual.values.tolist()}
test_pred = pd.DataFrame(d)
test_pred.head(5)
```

✓ 0.1s

	prediction	true_value
0	[5.070000171661377]	3.0
1	[4.510000228881836]	4.5
2	[2.7300000190734863]	3.5
3	[4.610000133514404]	4.0
4	[4.710000038146973]	3.5

It shows the deviation of the predicted rating and the true ratings for which the rmse value has been calculated over different embedding layers and epochs.

The proposed methodology gets the rmse value as




```
y_pred = np.round(model.predict([test_data['user_encoded'], test_data['movie_encoded']]), decimals=2)
y_actual = test_data['rating']
print(f'testing rms: {(mean_squared_error(y_actual, y_pred))}')
```

✓ 2.5s

```
testing rms: 0.16974757419414907
```

While comparing it with different other algorithms which we have tried to implement during our comprehensive analysis, we have drawn the conclusion using various other methods of recommendation system have tried to train and test our dataset on various other algorithms which include:

1. Content-Based Filtering using Cosine Similarity with dot product
2. Collaborative Filtering using Pearson correlation and Singular Value Decomposition
3. Hybrid RS is the conglomeration of Content-Based filtering and Collaborative Filtering using SVD using the weighted average method.

Results have been depicted in tabulated form

ALGORITHMS IMPLEMENTED	**RMSE SCORE**
CF using Pearson Correlation	1.1534
CF using MF SVD	1.0034
Hybrid Approach	0.9561
Proposed Algorithm	**0.1693**

The lesser the RMSE, the better the model

Recommended Movies based on proposed Algorithms is:

```
user_enter_id = int(np.random.choice(user_ID, 1))
print(f'recommendation for client: {user_enter_id}')
NCF_recommendation(model, user_enter_id, top_k=15)
```

3.7s

recommendation for client: 1342

movieId	prediction	title	genres
139644	5.585520	Sicario (2015)	Crime|Drama|Mystery
8376	5.551603	Napoleon Dynamite (2004)	Comedy
2020	5.353353	Dangerous Liaisons (1988)	Drama|Romance
5014	5.170085	I Am Sam (2001)	Drama
215	5.165090	Before Sunrise (1995)	Drama|Romance
3159	5.122891	Fantasia 2000 (1999)	Animation|Children|Musical|IMAX
628	5.099882	Primal Fear (1996)	Crime|Drama|Mystery|Thriller
66203	5.097860	He's Just Not That Into You (2009)	Comedy|Drama|Romance
1092	5.071863	Basic Instinct (1992)	Crime|Mystery|Thriller
27721	5.070566	Very Long Engagement, A (Un long dimanche de f...	Drama|Mystery|Romance|War
33004	5.066860	Hitchhiker's Guide to the Galaxy, The (2005)	Adventure|Comedy|Sci-Fi
134130	5.065013	The Martian (2015)	Adventure|Drama|Sci-Fi
1247	5.031280	Graduate, The (1967)	Comedy|Drama|Romance
88810	5.009164	Help, The (2011)	Drama
36	4.970178	Dead Man Walking (1995)	Crime|Drama

8. Conclusion & Future Scope

In this project, the recommendation system is improved using a meta-heuristic algorithm - The genetic Algorithm. NCF is used to predict the rating given by a user to a particular movie. Predicted movie ratings are then compared with actual ratings to find the accuracy of the NCF model. Then, in order to enhance movie rating prediction, a Genetic Algorithm is used to develop an optimum topology of an Artificial Neural Network. GA's major operators, like those of other EAs, are selection, crossover, and mutation. These techniques provide a method for programs to enhance their settings automatically. The new populations are created by repeatedly applying genetic operators to existing individuals in the population. Movie RS is enhanced with an RMSE score of 0.16 which is quite good as compared to other algorithms implemented.

Despite the fact that numerous attempts have been made in the past to develop recommendation system algorithms, there is still much room for improvement. Fine-tuning the edges, such as assigning high weight to current edges and low weight to older ones, and applying psychology to better manage the network's human features, can enhance the predictions. In the network, there may be duplicate, inactive, fake, or robotic nodes, which can be detected and weighed differently from active, genuine, and human nodes.

8. References

[1] Stitini, O., Kaloun, S., & Bencharef, O. (2022). An Improved Recommender System Solution to Mitigate the Over-Specialization Problem Using Genetic Algorithms. *Electronics*, *11*(2), 242.

[2] Rezaee, A. A., & Abravan, N. (2020, October). A hybrid friend-based recommendation system using the combination of Meta-heuristic Invasive weed and genetic algorithms. In *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 665-669). IEEE.

[3] Sarker, M. R. I., & Matin, A. (2021, February). A hybrid collaborative recommendation system based on matrix factorization and deep neural network. In *2021 international conference on information and communication technology for sustainable development (ICICT4SD)* (pp. 371-374). IEEE.

[4] Wei, G., Wu, Q., & Zhou, M. (2021). A hybrid probabilistic multiobjective evolutionary algorithm for commercial recommendation systems. *IEEE Transactions on Computational Social Systems*, *8*(3), 589-598.

[5] Katarya, R., & Verma, O. P. (2018). Recommender system with grey wolf optimizer and FCM. *Neural Computing and Applications*, *30*(5), 1679-1687.

[6] Färber, M., & Jatowt, A. (2020). Citation recommendation: approaches and datasets. *International Journal on Digital Libraries*, *21*(4), 375-405.

[7] Anjali, A., Sandhu, J. K., & Goyal, D. (2021, February). User Profiling in Travel Recommender System using Hybridization and Collaborative Method. In *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 143-148). IEEE.

[8] Zaremarjal, A. Y., & Yiltas-Kaplan, D. (2021, April). Semantic Collaborative Filtering Recommender System Using CNNs. In *2021 8th International Conference on Electrical and Electronics Engineering (ICEEE)* (pp. 254-258). IEEE.

[9] Khanse, S., Bhandari, P., Singru, R., Runwal, N., & Dharane, A. (2020, November). Comparative Study of Genetic Algorithm and Artificial Neural Network for Multi-class Classification based on Type-2 Diabetes Treatment Recommendation model. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 538-543). IEEE.

[10] Almaghrabi, M., & Chetty, G. (2020, November). Multilingual sentiment recommendation system based on multilayer convolutional neural networks (mcnn) and collaborative filtering based multistage deep neural network models (cfmdnn). In *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1-6). IEEE.

[11] Peng, Y., Hu, R., & Wen, Y. (2021, December). CA-NCF: A Category Assisted Neural Collaborative Filtering Approach for Personalized Recommendation. In *2021 IEEE International Conference on Progress in Informatics and Computing (PIC)* (pp. 241-247). IEEE.

[12] Yu, J., Shi, J., Chen, Y., Ji, D., Liu, W., Xie, Z., ... & Feng, X. (2021, March). Collaborative Filtering Recommendation with Fluctuations of User'Preference. In *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)* (pp. 222-226). IEEE.

[13] Li, B., Wan, S., Xia, H., & Qian, F. (2020, August). The Research for Recommendation System Based on Improved KNN Algorithm. In *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 796-798). IEEE.

[14] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor "Recommender Systems Handbook", Springer New York Dordrecht Heidelberg London.