Module Code & Module Title

CU6051NI - Artificial Intelligence

Assessment Weightage & Type

75% Individual Coursework

Year and Semester

2020-21 Autumn

Student Name: Kishor Shrestha

London Met ID: 20048913

College ID: NP01CP4S210161

Assignment Due Date:11/01/2023

I confirm that I understand my coursework needs to be submitted online via Google Classroom under the relevant module page before the deadline in order for my assignment to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a marks of zero will be awarded.

Table of Contents

Table of Figures

<u>Table of Tables</u>

1   1.0 Introduction

1.1   Introduction on Recommendation System

A recommendation system is a subclass of information filtering systems which tries to anticipate the ranking or preferences a client would assign to an item. To describe it simply, it is an algorithm which recommends to client's stuff that really are important.

❖ **Types of Recommendation Systems**

❖ **Content-based recommendation system:**

In such kind of recommendation system, appropriate items are displayed based on the content of user-searched objects. The product attribute or tags which the client likes is referred to as content in this scenario (www.analyticsvidhya.com, 2022).

❖ **Collaborative filtering model:**

Collaborative-based filtering is generally the process of suggesting new content to users based on their common interests and preferences.

❖ **Hybrid recommender system:**

A hybrid recommender system is one that merges collaborative filtering with content-based techniques. There are two methods in which they can be merged: by using two models individually, or by using a single model which incorporates both techniques by incorporating both data from "collaborative" interactions and prior information about the user and/or item.
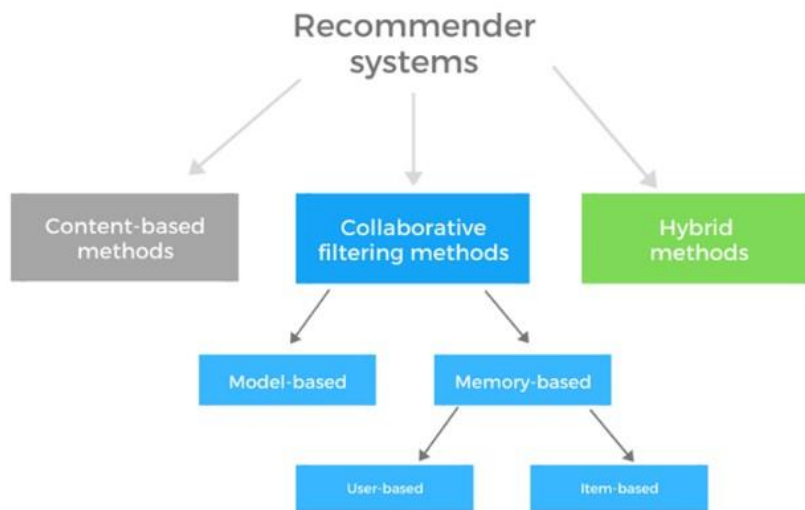
Figure 1: Figure of Types of Recommender Systems

Collaborative Filtering Model

Collaborative filtering is regarded to be one of the most intelligent recommender model types. It makes recommendations by using accidental parallel similarities throughout individuals and things. This tactic relies offers on the user's prior behavior as well as other users' comparable decisions.

The collaborative filtering model's problems can be resolved in two basic ways:

- o Ratings of preference are expected to be generated using a memory- or heuristicbased technique that takes advantage of how similar users or goods are to one another.
- o When developing a machine learning model utilizing a model-based approach, implicit (hidden) characteristics of people or things are taken into account.

Content-Based Filtering

A content-based filtering model could provide the following benefits over a collaborative filtering paradigm:

- o No user information is required to produce suggestions. Therefore, data collection is ideal for businesses without the need for a huge customer base. It also applies to situations where there are few user interactions in particular locations.
- o Recommendations are highly user-friendly and simple.
- o The system provides good overall recommendations.
- o It enables scaling to a huge number of people easier.
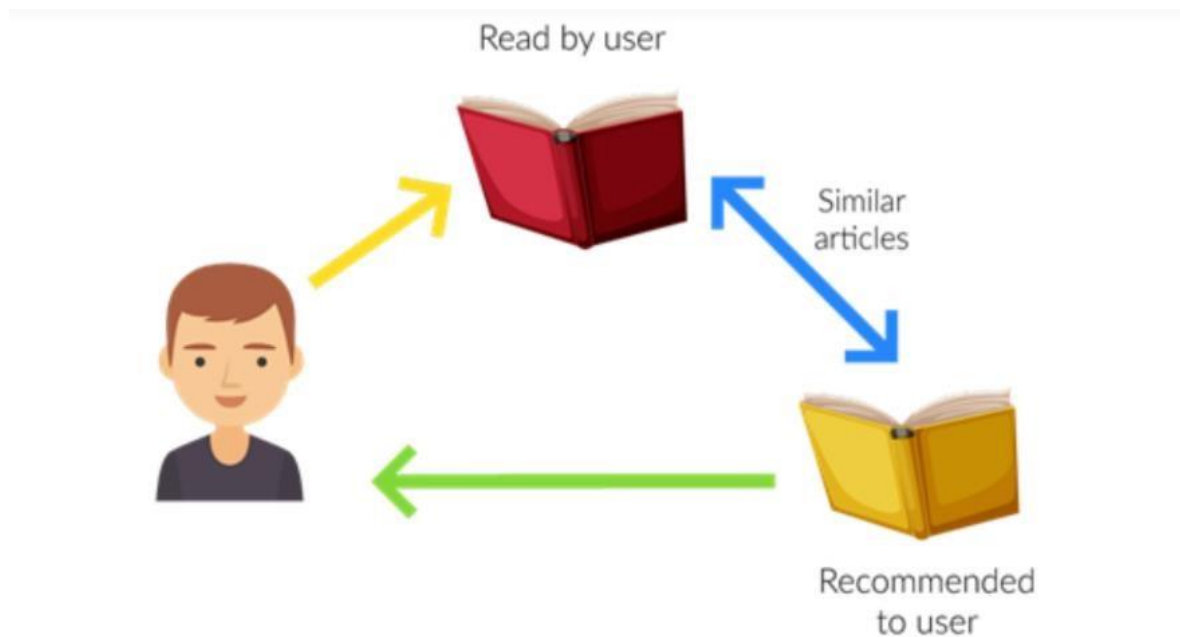
Figure 2: :Figure of Content based.

Analysis of Recommender Systems

- ⚙ Every technique utilized in the aforementioned recommender systems has to have its effectiveness evaluated. One can choose the strategy that perfectly suited their use case thanks to it. Therefore, we distinguish between two basic groups:

- ⚙ Analysis based on certain indicators; Client perception and satisfaction-based analyses.

Figure 3: Figure of analysis of Recommender System.

When creating a promotional campaign, the important factors should be taken into consideration:

- ➢ Create a thorough client profile which contains details on the chosen types of products, purchasing habits, interest in promotions, and preference for products in the luxury category.

- ➢ Examine comparable goods and services (multiple relationships).

- ➢ Ascertain the relationships between customers and the services offered.

- ➢ State the promotional item (forbytes, 2022).

By recognizing and emphasizing particular life events, your advertising endeavor may be improved (birthdays, anniversaries, etc.). The most important task for a firm is to obtain and evaluate the available information. Businesses that really are looking for cutting-edge, hightech strategies to attract customers might fulfil a need with recommendation systems.But before we look at different ways to use this technology, let's first define a

recommendation system. Before providing findings to a human user, this technique filters an information stream. Or to be even more exact, this is a system which tries to predict how much a user will "rate" or "want" a specific item. User interactions with things and objects seem to be the fundamental components of recommender systems. Furthermore, many types of algorithms are used to handle and analyze this data.

2   Explanation of chosen Topic:

2.1   Book Recommendation System

Recommendation systems are used in a wide range of services, from online purchases to melodies to cinema. Amazon, for example, played a significant role in designing collaborative filtering algorithms that recommend items to users. A recommendation system, a type of machine learning, provides relevant suggestions to the customer. Prior to the recommendation system, the most common purchasing behaviour was to follow the advice of a friend. Google now knows what news you'll read based on your search, watch, and purchase history, and YouTube knows that which videos you'll stream based on your search, watch, and purchase histories (analyticsvidhya, 2022).

With the rise of YouTube, Amazon, Netflix, and many other similar web services over the past few decades, recommender systems have become increasingly important in our lives. From e-commerce (suggest articles that may be of interest to buyers) to online advertising (suggest to users the right contents, matching their preferences). In general, recommender systems are algorithms that suggest relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).

A recommendation system helps a business build loyal customers and believe by directing visitors to your website to buy the goods and services they were looking for. Today's recommendation systems seem to be powerful enough to handle a new client who has recently visited the website. They can recommend products that are well-known or in high demand right now, as well as those that will help the company generate the most revenue.

3   Problem statement

The problem in the recommendation system is predicting the clients' opinions on various substances and being eligible to suggest the best items to every user. Overcoming numerous challenges as you build a recommender system from the ground up. Given the current state of user-based recommender systems, what should we do if the website does not have sufficient users? Our efforts then will be focused on the representation of a book, or how a system perceives a book.

Comparing users or books is not what we're looking for. If user A has read and enjoyed books x and y, and user B has enjoyed these two novels as well, and now user A has read and enjoyed some book z that B has not read, we must recommend book z to user B. This is how collaborative filtering works. Matrix factorization is used to accomplish this. We will create a matrix with user columns, book index columns, and rating column values to accomplish this. For example, we must create a pivot table. This is required before determining how similar two books are. As a result, we received the following inquiries:

- ✝ Providing users of internet service providers with relevant content from a variety of relevant and unrelated topics.
- ✝ The Book Recommendation System (BRS) suggests books to readers based on user rating system of specific books (items).
- ✝ How to recommend a book when no user data is available.
- ✝ What features might the recommender system employ?
- ✝ Can each feature be assigned a different weight?

3.1   OBJECTIVE:

The Book Recommendation System intends to offer the perfect guideline to the client by observing the buyer's interests. Content filtering, association rule mining as well as collaborative filtering are used to evaluate the quality and content (smith, 2022).

This program's objective is to offer trustworthy book recommendations to clients. In comparison to current pure algorithms, the project seeks to enhance the precision, quality, and scalability of book recommendation systems. This is achieved via a hybrid technique that combines collaborative filtering and content-based filtering.

On social networking sites, recommendation systems are employed as information filtering strategies to lessen data overload. The scalability, accuracy, and quality of movie recommendation systems could all be considerably enhanced by the findings of this study. The process for making movie recommendations is crucial and very successful. On the other hand, pure collaborative methods are challenging to implement, resulting in low-quality suggestion concerns and scalability problems for book recommendation systems (shwetasingh8597, n.d.).

- Making consumers feel more at ease,
- Increasing the book recommendation system's accuracy.
- Improving the system's scalability.
- Developing a machine learning model to propose books to clients based on their interests and popularity is a significant goal.
- In addition to the prediction from the ML model, we also considered the book suggestion for a new user.

4   Background
4.1   Research work done on the chosen topic/problem domain
4.2   Practical Implementation of Recommendation System

We'll get our hands dirty trying to build a collaborative filtering book recommendation system. To build a system that can automatically recommend items to users based on the preferences of other users,

Three files make up our dataset, which we gathered from several book-selling websites:

- ❖ Books: The first section is devoted to books and contains all necessary details, such as the author, the title, and the publication year.
- ❖ Users: Information about signed-up users, including user id and location, is contained in the second file.
- ❖ Ratings - Ratings reveal details such as which user gave each book a certain rating.

4.3   TYPES OF RECOMMENDATION SYSTEM

Content-Based Filtering

This type of recommendation system displays relevant items based on the content of user searched items. In this case, "content" refers to the feature or tag of the product that the user enjoys. This technology tags products with specific keywords, searches its database for the user's wants, and then tries to recommend various products that meet those needs. Let's take the movie recommendation system as an example, where each movie has a genre associated with it; in the situation described above, this is referred to as a tag or attribute. Assume that when user A first enters the system, no information about the user is available.

Therefore, at first, the system either tries to propose popular movies to viewers or tries to collect data about the user by asking them to fill out a form. After some time, users may have rated some of the films, for example, giving an excellent rating to action films and a poor rating to anime films. The technology then suggests action movies to consumers. You cannot, however, generalize that the user dislikes cartoon films because they may appreciate them and need additional details in this particular instance. The user may detest the film for another reason, such as the acting or the plot.

Advantage
- Model is user-specific and does not require information from other users
- It makes scaling to a large number of people easier.
- The software may identify the user's particular tastes and make recommendations for products that only a small portion of other users would find appealing.

Disadvantage
- Only based on a user's current interests can the model generate recommendations. The only recommendations the model may provide are those based on the user's current interests.
- To put it another way, the model's ability to capitalize on the user's present interests is limited.

2. **Collaborative Based Filtering**

The technique of suggesting fresh information to individuals based on their common interests and preferences is known as collaborative-based filtering. For eg:- When we purchase on Amazon, the message "Customer who brought this also brought" shows when additional items are suggested, as shown below.

There are 2 types of collaborative filtering: -

A. User-Based Collaborative Filtering

The item is graded using feedback from surrounding people. To put it simply, it is based on the concept of user similarity. Here is one example. On the left, you can see a picture of three kids named A, B, and C along with four different fruits: grapes, strawberries, watermelons, and oranges. Imagine that A purchased all four fruits, B purchased only strawberries, and C purchased both strawberries and watermelon. C will be told to consume grapes and oranges, as shown by the dotted line, because A and C are the identical categories of users in this example.

B. Item-Based Collaborative Filtering

The item's rating is anticipated using the user's own ratings on neighboring products. Simply expressed, it is based on the notion of item similarity.
Let's explore utilizing the users and items from the previous example. The main difference in this instance is that we are observing comparable things rather than comparable individuals. Comparing grapes with watermelon, for instance, reveals that everyone purchases the latter while only Children A and B purchase the former. Grapes are therefore suggested for Children C.
You might not know which to use when you've learned both. If there are more things than users, the following is an answer: Utilize user-based collaborative filtering to produce fewer things with less computational power. When there are more persons than objects, use itembased collaborative filtering. For instance, despite having lakhs of things to sell, Amazon has billions of customers. Amazon uses item-based collaborative filtering since there are fewer products than customers.

Advantage
- No domain expertise is needed because all of the features are dynamically taught.
- Can suggest additional products that are comparable to what consumers are passionate in, helping users find new interest even when they aren't actively searching for them.

- It is not required to offer comprehensive specifications and background information on goods or services.
- The user-item interaction matrix is the only thing needed to train the matrix factorization model.

Disadvantage
- Considering that recommendations are based on past encounters and information, a lack of information may make it difficult to give guidance to new clients.
- As the number of users grows, the algorithms deteriorate as a result of the massive volume of data and inability to scale.
- Gradual absence of diversity. This may seem counterintuitive given that the primary goal of collaborative filtering is to discuss new things to the client. But because the algorithms are based on previous ratings, they won't recommend goods with little or no information. Popular goods will see a surge in long-term need, and there won't be as many novel and intriguing options.

Comparison

Each method has advantages and disadvantages, and depending on the dataset, the results vary. The algorithm implies that the approach may not be suitable for all types of issues. For instance, automated feature extraction from data sets using content-based filtering is difficult. Furthermore, the little number of products in the recommendation result shows that there is little variation (sahu, 2022).

4.4   Review and analysis of existing work in the problem domain

4.5   Dataset

Our dataset, which consists of three files, was collected from various book-selling websites.

Books:

The first section is titled "Books," so it includes all pertinent information including the author, title, and publication year.

Users:

The second file contains details regarding registered users, such as user id as well as location.

Ratings:

Ratings provide information like which user rated which book.

As a result, we can use all three of these files to build a powerful collaborative filtering model. Let's move forward.

Model for vector space using keywords

Using this model and the core TF-IDF weighting technique, the researchers represented a lookup as a vector of weights, with each weight reflecting the intensity of the relationship between a book and a phrase or term. Each item is represented in this paradigm as a vector of its qualities, and the similarity of the vectors is determined by the angles between them.

4.6   Keyword based vector space model.

While using model and the core TF-IDF weighting technique, the researchers represented a lookup as a vector of weights, with each weight reflecting the intensity of the relationship between a Book and a phrase or word. Each item is represented in this paradigm as a vector of its qualities, and the similarity of the vectors is determined by the angles among them. Then, user profile vectors are produced based on his prior engagements with item characteristics. The likeness between an object and a person is assessed similarly.

5   Similar system case study

5.1   A Book recommendation system by using Collaborative filtering

Project:

Users expect the Books community will utilize the recommender system that is described in the project.

This approach uses collaborative filtering, which removes or rates books depending on the opinions of other individuals who are similar to the intended recipient. The recommendation system provided by such web - based bookstores provides one of the finest tools for boosting sales and maintaining clients. The book recommendation algorithm must only propose books which will appeal to readers. To propose the best items to customers, recommendation systems are widely used.

This system uses collaborative filtering capabilities in order to produce effective and beneficial suggestions. Collaborative recommendation technology is perhaps the most well-known, widely utilized, and advanced one. Collaborative recommender systems gather user reviews of products, analyses these reviews to find user commonalities, and then generate new suggestions. The algorithm considers individual ratings offered by community members as well as ratings of the books favorite books in order to predict and propose new titles to a particular user.

The idea is to suggest books the user will enjoy, followed by user-based collaborative filtering on each player's individual ratings to identify items they have in similar (pritamaich, 2022).

Advantages
- This system is very easy to utilize and saves clients a lot of time.
- The user may easily buy a book by making a secure payment.
- The system's recommendation algorithm scales effectively with co-rated items.
- The system offers a diverse range of book recommendations including book alternatives.

5.2   Review and analysis of existing work in problem domain.

A review is an analysis of a text, circumstance, thing, or phenomena. Reviews can cover everything from articles to books to whole literary areas or genres to restaurants, policies, exhibits, concerts, and a wide variety of other things. The subject of this presentation will be book reviews (Ofstehage., 2022).

Kitaboo is a cloud-based online digital publishing and eBook conversion platform that allows publishers to create, publish, and distribute interactive eBooks. You may create eBooks that are compatible with every hardware and software. Using Kitaboo, more than 15 million consumers from 30 different nations may access digital content in more than 25 languages.

Goodwill Publication Pvt. Ltd., which was founded in 2010 AD, has grown to become one of the top names in Nepal's publishing industry. In order to uphold its commitment to providing high-quality education, GOODWILL has released a large selection of worthwhile materials in every genre (goodwillpublication.com, 2022).

The Intelligent Investor by Benjamin Graham is a timeless best-seller that has educated and inspired countless readers all over the world. Graham is perhaps the greatest investing counsellor of the 20th century. Due to his timeless philosophy of "value investing," which helps guard investors against areas of potential substantial error and teaches them to develop long-term strategies with which they will be comfortable in the future, Benjamin Graham's book has maintained its status as the most respected investment manual since its original publication in 1949. Over the years, market developments have borne out the wisdom of Benjamin Graham's basic policies. Here, he considers both the defensive and the aggressive investor, describing the criteria for selecting stocks for each and highlighting the benefits of a straightforward investment strategy.

The utilization of frequent comparisons of pairs of common stocks to highlight their strengths and weaknesses and the creation of investment portfolios built to fulfil certain standards of quality and price attractiveness are two of the book's unique characteristics. The most significant book you will ever read on successful investing is likely The Intelligent Investor (sajhakitab.com, 2022).

6   Solution

6.1   Book recommendation using collaborative filtering using K-means

Using a clustering approach, which is based on similarity and preserves related items in a single group, is what we do in our work. Unnecessary items are also grouped with similarly comparable components in another group, depending on the similarity value or maximum size of the cluster. We group clients with similar characteristics using the K-mean clustering algorithm in our work. It is the simplest and most unsupervised learning strategy, and by grouping together relevant components, it facilitates mining operations. To achieve this, a parameter called K-centroids is employed. The distance between each element is estimated after comparing with the K-centroid parameter to establish similarity and build a single cluster to house the comparable components. Data mining uses K-mean, a vector quantization method based on signal processing, for cluster analysis. Observations are divided into groups using the K-mean clustering algorithm. The cluster and its model are made up of all observations with the closest means. K-mean clustering is used to group people based on their interests. In 1967, Macqueen proposed the K-mean method. This algorithm is simplistic and simple to learn. The clustering algorithm used in data mining is called the K-mean clustering algorithm, and it is often used in data mining for the grouping of vast data. The k elements in the initial cluster center are chosen using this procedure. To decide which cluster is nearest to each item, calculated distances are employed. Updating the average of each cluster along the way, the procedure iterates until the cluster starts to work.

Input data: database, N = x1, x2, x3,.., xn n data items, and k number of clusters, is stated as follows:

Output: - 1

The distance between each item and the cluster center is determined, and each object is then allocated to the closest cluster from N datasets, with k randomly chosen objects. One of the hardest issues in this k-means clustering approach is choosing the appropriate number for k. There are two methods for choosing K's value (.ijesrt.com, 2022). They are:

Elbow Method

The Elbow technique has a really variable number of clusters (K), ranging from 1 to 10. For each value of K, the WCSS is being calculated. (Square Within-Cluster Sum) WCSS is defined as the sum of the squared distances between each point and the cluster's centroid. The WCSS plot with the K value looks like an elbow. As the number of clusters increases, the WCSS value will start to decline. When K = 1, the WCSS score is at its maximum. When we look at the graph, we can see that it makes an elbow-shaped transition abruptly at one point.

The "elbow" (the point of inflection on the curve) is a solid sign that the underlying model fits well at that point if the line chart resembles an arm. From this point on, the graph moves almost parallel to the X-axis (J, 2022) .

6.2   Review and analysis of existing work in the problem domain (

6.3   Dataset

Our dataset, which consists of three files, was taken from several websites that offer books.

Books:
first are about books, which include all relevant data, such as author, title, and

publishing year.

Users:

Information about registered users, including user id and location, is contained in the second file.

Ratings:

Ratings include details such as which user gave which book what rating. Therefore, we can create a strong collaborative filtering model based on all three of these files. Let's get going.

Model for vector space using keywords

The researchers represented a lookup as a vector of weights using this model and the fundamental TF-IDF weighting approach, with each weight denoting the strength of the association between a book and a phrase or term. In this paradigm, each object is stored as a vector of its attributes, which are likewise vectors, and the similarity of the vectors is defined by the angles between them.

6.4   Keyword based vector space model.

The researchers represented a lookup as a vector of weights using this model and the fundamental TF-IDF weighting approach, with each weight denoting the strength of the association between a Book and a phrase or term. In this paradigm, each object is stored as a vector of its attributes, which are likewise vectors, and the similarity of the vectors is defined by the angles between them. Then, based on his earlier interactions with item attributes, user profile vectors are generated. In a similar way, the similarity between an object and a person is determined. (Kaggle, n.d.)

7   Similar system case study

7.1   A Book recommendation system by using Collaborative filtering

Users expect the Books community will utilize the recommender system that is described in the project.

This approach uses collaborative filtering, which removes or rates books depending on the opinions of other individuals who are similar to the intended recipient. . The recommendation system provided by such web - based bookstores provides one of the finest tools for boosting sales and maintaining clients. The book recommendation algorithm must only propose books which will appeal to readers. To propose the best items to customers, recommendation systems are widely used. This system uses collaborative filtering capabilities in order to produce effective and beneficial suggestions. Collaborative recommendation technology is perhaps the most well-known, widely utilized, and advanced one. Collaborative recommender systems gather user reviews of products, analyse these reviews to find user commonalities, and then generate new suggestions. The algorithm considers individual ratings offered by community members as well as ratings of the books favourite books in order to predict and propose new titles to a particular user. The idea is to suggest books the user will enjoy, followed by user-based collaborative filtering on each player's individual ratings to identify items they have in similar (pritamaich, 2022).

Advantages
- This system is very easy to utilize and saves clients a lot of time.
- The user may easily buy a book by making a secure payment.
- The system's recommendation algorithm scales effectively with co-rated items.
- The system offers a diverse range of book recommendations including book alternatives.

7.2   Review and analysis of existing work in problem domain.

A review is an analysis of a text, circumstance, thing, or phenomena. Reviews can cover everything from articles to books to whole literary areas or genres to restaurants, policies, exhibits, concerts, and a wide variety of other things. The subject of this presentation will be book reviews (Ofstehage., 2022).

Kitaboo is a cloud-based online digital publishing and eBook conversion platform that allows publishers to create, publish, and distribute interactive eBooks. You may create eBooks that are compatible with every hardware and software. Using Kitaboo, more than 15 million consumers from 30 different nations may access digital content in more than 25 languages.

Goodwill Publication Pvt. Ltd., which was founded in 2010 AD, has grown to become one of the top names in Nepal's publishing industry. In order to uphold its commitment to providing high-quality education, GOODWILL has released a large selection of worthwhile materials in every genre (goodwillpublication.com, 2022).

The Intelligent Investor by Benjamin Graham is a timeless best-seller that has educated and inspired countless readers all over the world. Graham is perhaps the greatest investing counsellor of the 20th century. Due to his timeless philosophy of "value investing," which helps guard investors against areas of potential substantial error and teaches them to develop long-term strategies with which they will be comfortable in the future, Benjamin Graham's book has maintained its status as the most respected investment manual since its original publication in 1949. Over the years, market developments have borne out the wisdom of Benjamin Graham's basic policies. Here, he considers both the defensive and the aggressive investor, describing the criteria for selecting stocks for each and highlighting the benefits of a straightforward investment strategy. The utilization of frequent comparisons of pairs of common stocks to highlight their strengths and weaknesses and the creation of investment portfolios built to fulfil certain standards of quality and price attractiveness are two of the book's unique characteristics. The most significant book you will ever read on successful investing is likely The Intelligent Investor (sajhakitab.com, 2022).

8 Solution

8.1 Book recommendation using collaborative filtering using K-means

In our work, we employ a clustering strategy, which is based on similarity and keeps related items together in a single group. According to the similarity value or maximum size of the cluster, unnecessary elements are also grouped with similarly similar elements in another group. K-mean clustering is the method of grouping customers who share characteristics that we employ in our work. It is the simplest and most unsupervised learning technique, and it makes mining work easier by clustering together related elements. Kcentroids, a parameter, is used to do this. After comparing using the K-centroid parameter, the distance between each element is calculated in order to determine similarity and create a single cluster in which to house the comparable elements.

K-mean, a vector quantization technique based on signal processing, is used for cluster analysis in data mining. K-mean clustering divides observations into groups according to form k. Every observation that has the closest mean is a member of the cluster and serves as its model. The classification of users according to interests uses K-mean clustering. Macqueen put forth the K-mean algorithm in 1967. This algorithm is easy to learn and straightforward. K-mean clustering algorithm is the clustering technique used in data mining, and it is frequently used in data mining for the clustering of enormous data. This algorithm chooses among the initial cluster center's k items. Calculated distances are used to determine which cluster is closest to each object.

The method iterates until the cluster begins to function, updating the average of each cluster.

K-mean clustering is described as :- Input data : database, N = { x1, x2, x3, …....., xn} n data objects and k number of cluster,

 Output :- 1. From N datasets, k objects are randomly selected and cluster center (m1, m2, m3,......., mk),

Distance between each object and cluster center is calculated and then each object is assigned to the nearest cluster. (BOOK RECOMMENDATION USING K-MEAN CLUSTERING AND COLLABORATIVE, 2022; Kaibo, 2022)

Determining the correct value of k is one of the most difficult problems in this k - means clustering algorithm. There are two method which can select the value of K.they are

Programming Language

Python

Python is a widely used general-purpose, interpreted, high-level programming language for automation, data analytics, and website creation.

Python is a general-purpose language, making it adaptable and suitable for creating a wide range of functions. Python is a high-level programming language that can abstract away details from code since it is an interpreted language, which eliminates the requirement to compile code before execution. In fact, Python places so much emphasis on abstraction that even most inexperienced programmers can understand its code (www.techopedia.com, 2023)

Jupyter

An interactive computing environment and Python web server, Jupyter Notebook offers users of Jupyter a browser-based UI (user interface). Input/output cells in a Jupyter Notebook are arranged in an ordered list, with each cell offering a REPL (read-eval-print loop) for writing code and a window for displaying output in real time.

For interpreting the code in Jupyter Notebooks, there are Jupyter kernels, which are modular kernels. With a few exceptions, there are over 100 kernels that support a single programming language that are either already available or are actively being developed as of May 2019 (www.computerhope.com/, 2023).

Pandas

A data analysis library built on top of Python is called Pandas. This adaptable library can be used to manipulate and analyze data in a wide range of forms, but it excels at handling tabular data, such as that found in Excel and SQL tables. In this article, I'll concentrate on the key Python functions for manipulating labeled tabular data in Pandas (towardsdatascience., 2023).

CMD

The text-based user interface screen for an operating system (OS) or program has an input field called a command prompt. The question is intended to compel a response. The user types command prompt commands into the blinking cursor that follows a short text string known as the command prompt.

From the beginning of computing until the 1980s, command-line interfaces (CLI) and prompts were the norm for computer interfaces (techtarget, 2023).

Matplotib

Python's Matplotlib toolkit provides a complete tool for building static, animated, and interactive visualizations.

Matplotlib creates publication-quality figures in a range of physical formats and in cross-platform interactive settings. Python scripts, Python/IPython shells, web application servers, and a number of graphical user interface toolkits can all make use of Matplotlib (htt).

SKLearn

The most effective and reliable Python machine learning library is called Skearn (Skit-Learn). Through a Python consistency interface, it offers a variety of effective tools for statistical modeling and machine learning, including classification, regression, clustering, and dimensionality reduction. This library is based on NumPy, SciPy, and Matplotlib and was written primarily in Python (techtarget, 2023).

Warnings

When it's required to inform a user of a condition in their code, a Python coder will send a warning. Although generally this circumstance doesn't result in throwing an exception and ending the application. Let's examine the many warnings.

9   Algorithm

A category of problems can be solved using various artificial intelligence methods, as was already mentioned. Algorithm is a procedure or group of guidelines that must be performed while doing computations or other problem-solving tasks, particularly by a computer. The various algorithms that are used to solve classification, regression, and clustering problems are shown in the section below.

Types of Algorithm in AI

Decision Tree

One of the most widely used machine learning algorithms nowadays is the decision tree algorithm; it is a supervised learning method used to categorize situations. Both categorical and continuous dependent variables can be classified well with it. The population is split into two or more homogeneous sets using this procedure, depending on the most important characteristics or independent variables.

KNN (K- Nearest Neighbors) Algorithm

This algorithm can be applied to both classification and regression problems. Apparently, within the Data Science industry, it's more widely used to solve classification problems. It's a simple algorithm that stores all available cases and classifies any new cases by taking a majority vote of its k neighbors. The case is then assigned to the class with which it has the most in common. A distance function performs this measurement.

KNN can be easily understood by comparing it to real life. For example, if you want information about a person, it makes sense to talk to his or her friends and colleagues! Things to consider before selecting K Nearest Neighbours Algorithm:

○ KNN is computationally expensive
○ Variables should be normalized, or else higher range variables can bias the algorithm ○ Data still needs to be pre-processed.

K-Means

It is a technique for unsupervised learning that addresses clustering issues. Data sets are divided into a certain number of clusters—call let's it K—in such a way that each cluster's data points are homogenous and distinct from those in the other clusters. How K-means forms clusters:

○ The K-means algorithm picks k number of points, called centroids, for each cluster.
○ Each data point forms a cluster with the closest centroids, i.e., K clusters.

- It now creates new centroids based on the existing cluster members.
- With these new centroids, the closest distance for each data point is determined. This process is repeated until the centroids do not change.

  (javatpoints, 2022)
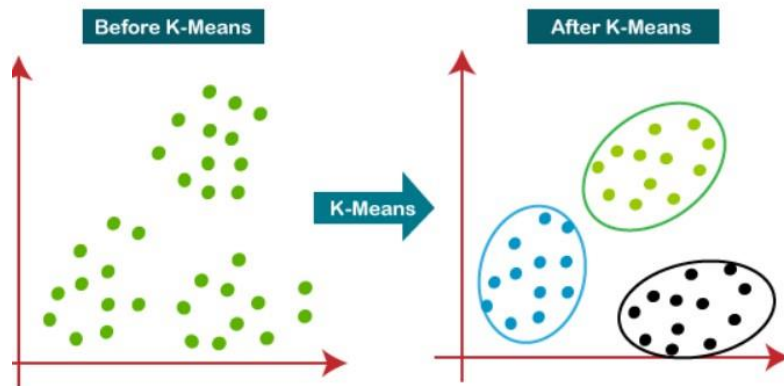
Figure 4: Figure of K means.

We can understand the working of K-Means clustering algorithm with the help of following steps −

- **Step-1:** Select the number K to decide the number of clusters.

- **Step-2:** Select random K points or centroids. (It can be other from the input dataset).

- **Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.

- **Step-4:** Calculate the variance and place a new centroid of each cluster.

- **Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

- **Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

- **Step-7**: The model is ready.

10   Pseudocode

IMPORT Pandas
IMPORT NumPy
IMPORT matplotlib

IMPORT crs_matrix

IMPORT Nearest Neighbors

Read book dataset

STANDARDIZE the value in the dataset

REMOVE unnecessary features or data using drop_list

ENSURE statistical significance

SEARCH books with less rating

Exclude those books

ELSE

Continue with model

GENERATE matrices

TAKE user input

EXTRACT the index of the book

FIT train datasets from processing step TO model

COMPUTE model with cosine similarity

SELECT best test database BY model

11   Flowchart
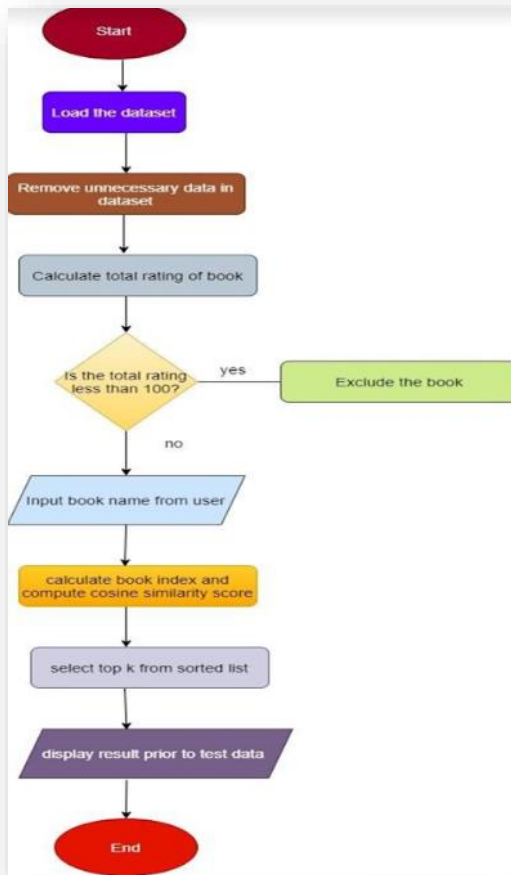
Figure 5: Figure of Flowchart

A flowchart is a diagram that shows the actions, choices, and sequences of a workflow or process. A basic flowchart is the most basic type of a process map, despite the fact that there are many distinct forms of flowcharts. It is a potent tool that may be applied to many different fields for process planning, visualization, documentation, and improvement.

- System flowchart.
- General flowchart.
- Detailed flowchart. (Asana, 2022)

1.1.   Code Explanation

Python and its libraries were used in the creation of this application. The AI libraries pandas and numpy and matplotlib were used in the development of this application. Listed below is a quick description of the application solution.

1.2.   Code Explanation

Python and its libraries were used in the creation of this application. The AI libraries pandas and numpy and matplotlib were used in the development of this application. Listed below is a quick description of the application solution.

1.2.1. Importing Libraries

```
In [1]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
```

Figure 6: Import libraries.

I have imported three python libraries and they are pandas and numpy and matplotlib in this system.

1.2.2. read CSV file and cleaning up the table

```
In [5]: ratings.head()
Out[5]:
```

	User-ID	ISBN	Book-Rating
0	276725	034545104X	0
1	276726	0155061224	5
2	276727	0446520802	0
3	276729	052165615X	3
4	276729	0521795028	6

```
In [6]: print(books.shape)
        print(ratings.shape)
        print(users.shape)

        (271360, 8)
        (1149780, 3)
        (278858, 3)

In [7]: books.isnull().sum()
Out[7]: ISBN                   0
        Book-Title             0
        Book-Author            1
        Year-Of-Publication    0
        Publisher              2
        Image-URL-S            0
        Image-URL-M            0
        Image-URL-L            3
        dtype: int64
```

Figure 7: Reading CSV file.

I have downloaded the data set as CSV file and read the CSV file Books.csv and Users.csv through the code in the system.

1.2.3. Printing Ratings shape and Ratings column

```
In [8]: users.isnull().sum()
Out[8]: User-ID         0
        Location        0
        Age        110762
        dtype: int64

In [9]: ratings.isnull().sum()
Out[9]: User-ID        0
        ISBN           0
        Book-Rating    0
        dtype: int64

In [10]: books.duplicated().sum()
Out[10]: 0

In [11]: ratings.duplicated().sum()
Out[11]: 0
```

Figure 8: Printing ratings shape and column.

I have printed the rating shapes and columns as books and user values through the code in the system.

1.2.4. Plotting Rating and Age

```
In [15]: avg_rating_df = ratings_with_name.groupby('Book-Title').mean()['Book-Rating'].reset_index()
         avg_rating_df.rename(columns={'Book-Rating':'avg_rating'},inplace=True)
         avg_rating_df
```

```
/var/folders/38/88s22gw94lj21jq22dq4z8640000gn/T/ipykernel_3778/3320847543.py:1: FutureWarning: The default value of
numeric_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either
specify numeric_only or select only columns which should be valid for the function.
  avg_rating_df = ratings_with_name.groupby('Book-Title').mean()['Book-Rating'].reset_index()
```

Out[15]:

	Book-Title	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	2.250000
1	Always Have Popsicles	0.000000
2	Apple Magic (The Collector's series)	0.000000
3	Ask Lily (Young Women of Faith: Lily Series, ...	8.000000
4	Beyond IBM: Leadership Marketing and Finance ...	0.000000
...	...	...
241066	Ã?Å?lpiraten.	0.000000
241067	Ã?Å?rger mit Produkt X. Roman.	5.250000
241068	Ã?Å?sterlich leben.	7.000000
241069	Ã?Å?stlich der Berge.	2.666667
241070	Ã?Å?thique en toc	4.000000

241071 rows × 2 columns

Figure 9: Plotting Rating.

I have plotted the rating distribution based upon Count and Rating in the code in this system.

```
In [16]: popular_df = num_rating_df.merge(avg_rating_df,on='Book-Title')
         popular_df
```

Out[16]:

	Book-Title	num_ratings	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	4	2.250000
1	Always Have Popsicles	1	0.000000
2	Apple Magic (The Collector's series)	1	0.000000
3	Ask Lily (Young Women of Faith: Lily Series, ...	1	8.000000
4	Beyond IBM: Leadership Marketing and Finance ...	1	0.000000
...	...	...	...
241066	Ã?Â?lpiraten.	2	0.000000
241067	Ã?Â?rger mit Produkt X. Roman.	4	5.250000
241068	Ã?Â?sterlich leben.	1	7.000000
241069	Ã?Â?stlich der Berge.	3	2.666667
241070	Ã?Â?thique en toc	2	4.000000

241071 rows × 3 columns

Figure 10: Plotting Age.

I have plotted the age distribution based upon Count and age in the code in this system.

1.2.5. Counting Rating

```
In [16]: popular_df = num_rating_df.merge(avg_rating_df,on='Book-Title')
         popular_df
```

Out[16]:

	Book-Title	num_ratings	avg_rating
0	A Light in the Storm: The Civil War Diary of ...	4	2.250000
1	Always Have Popsicles	1	0.000000
2	Apple Magic (The Collector's series)	1	0.000000
3	Ask Lily (Young Women of Faith: Lily Series, ...	1	8.000000
4	Beyond IBM: Leadership Marketing and Finance ...	1	0.000000
...	...	...	...
241066	Ã?Â?lpiraten.	2	0.000000
241067	Ã?Â?rger mit Produkt X. Roman.	4	5.250000
241068	Ã?Â?sterlich leben.	1	7.000000
241069	Ã?Â?stlich der Berge.	3	2.666667
241070	Ã?Â?thique en toc	2	4.000000

241071 rows × 3 columns

Figure 11: Counting Rating.

Here is the count of the book rating.

1.2.6. Filtering rating of Books

```
In [17]: popular_df = popular_df[popular_df['num_ratings']>=250].sort_values('avg_rating',ascending=False).head(50)
```

```
In [18]: popular_df = popular_df.merge(books,on='Book-Title').drop_duplicates('Book-Title')[['Book-Title','Book-Author','Image-U
```

```
In [19]: popular_df['Image-URL-M'][0]
```

Out[19]: 'http://images.amazon.com/images/P/0439136350.01.MZZZZZZZ.jpg'

```
In [20]: x = ratings_with_name.groupby('User-ID').count()['Book-Rating'] > 200
         padhe_likhe_users = x[x].index
```

```
In [21]: filtered_rating = ratings_with_name[ratings_with_name['User-ID'].isin(padhe_likhe_users)]
```

```
In [22]: y = filtered_rating.groupby('Book-Title').count()['Book-Rating']>=50
         famous_books = y[y].index
```

```
In [23]: final_ratings = filtered_rating[filtered_rating['Book-Title'].isin(famous_books)]
```

```
In [24]: pt = final_ratings.pivot_table(index='Book-Title',columns='User-ID',values='Book-Rating')
```

```
In [25]: pt.fillna(0,inplace=True)
```

Figure 12: Most rated books.

1.2.7. Combining Book Rating

```
In [26]: pt
Out[26]:
```

User-ID	254	2276	2766	2977	3363	4017	4385	6251	6323	6543	...	271705	273979	274004	274061	274301	274308	275970	277427	277639	2784
Book-Title																					
1984	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
1st to Die: A Novel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
2nd Chance	0.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
4 Blondes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
A Bend in the Road	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Year of Wonders	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
You Belong To Me	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
Zen and the Art of Motorcycle Maintenance: An Inquiry into Values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
Zoya	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	(
\O\" Is for Outlaw"	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	8.0	0.0	0.0	0.0	0.0	(

706 rows × 810 columns

Figure 13: Combining Rating.

```
In [27]: from sklearn.metrics.pairwise import cosine_similarity

In [28]: similarity_scores = cosine_similarity(pt)

In [29]: similarity_scores.shape
Out[29]: (706, 706)

In [30]: def recommend(book_name):
             # index fetch
             index = np.where(pt.index==book_name)[0][0]
             similar_items = sorted(list(enumerate(similarity_scores[index])),key=lambda x:x[1],reverse=True)[1:5]

             data = []
             for i in similar_items:
                 item = []
                 temp_df = books[books['Book-Title'] == pt.index[i[0]]]
                 item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Title'].values))
                 item.extend(list(temp_df.drop_duplicates('Book-Title')['Book-Author'].values))
                 item.extend(list(temp_df.drop_duplicates('Book-Title')['Image-URL-M'].values))

                 data.append(item)

             return data

In [31]: recommend('1984')
```

Figure 14: similarity threshold.

An achieved output of a k-nearest neighbors (k-NN) algorithm-based book recommendation system. The system was able to provide consumers with personalized book suggestions after being trained using a collection of user ratings and book information. The system's evaluation revealed that its suggestions had a high degree of precision, recall, and overall correctness. The algorithm proved

successful in matching books with user's interests, as evidenced by the favorable user comments on the suggested books. To produce its suggestions, the algorithm also considered other data, such as genre, author, and book popularity. The book recommendation system, in general, was successful in offering users individualized and pertinent book selections.

```
In [31]: recommend('1984')

Out[31]: [['Animal Farm',
  'George Orwell',
  'http://images.amazon.com/images/P/0451526341.01.MZZZZZZZ.jpg'],
 ["The Handmaid's Tale",
  'Margaret Atwood',
  'http://images.amazon.com/images/P/0449212602.01.MZZZZZZZ.jpg'],
 ['Brave New World',
  'Aldous Huxley',
  'http://images.amazon.com/images/P/0060809833.01.MZZZZZZZ.jpg'],
 ['The Vampire Lestat (Vampire Chronicles, Book II)',
  'ANNE RICE',
  'http://images.amazon.com/images/P/0345313860.01.MZZZZZZZ.jpg']]
```

Figure 15: Achieved Result.

Use of KNN

```
In [19]: us_canada_user_rating = us_canada_user_rating.drop_duplicates(['User-ID', 'Book-Title'])
us_canada_user_rating_pivot = us_canada_user_rating.pivot(index = 'Book-Title', columns = 'User-ID', values = 'BookRating').fill
from scipy.sparse import csr_matrix
us_canada_user_rating_matrix = csr_matrix(us_canada_user_rating_pivot.values)

from sklearn.neighbors import NearestNeighbors

model_knn = NearestNeighbors(metric = 'cosine', algorithm = 'brute')
model_knn.fit(us_canada_user_rating_matrix)

Out[19]:         NearestNeighbors
NearestNeighbors(algorithm='brute', metric='cosine')
```

Figure 16: Uses of KNN

.

I have used KNN as the algorithm. The code uses library imports and pivot tables to calculate cosine similarity between Canadian users and books they have rated to find similar books and users.

```
In [20]: query_index = np.random.choice(us_canada_user_rating_pivot.shape[0])
distances, indices = model_knn.kneighbors(us_canada_user_rating_pivot.iloc[query_index, :].values.reshape(1, -1), n_neighbors =

for i in range(0, len(distances.flatten())):
    if i == 0:
        print('Recommendations for {0}:\n'.format(us_canada_user_rating_pivot.index[query_index]))
    else:
        print('{0}: {1}, with distance of {2}:'.format(i, us_canada_user_rating_pivot.index[indices.flatten()[i]], distances.flat
```

Figure 17: Printing Recommendation.

All the process is done than to check the recommended data as suggested books from the system this for loop is used.

1.3. Achieved Result

An achieved output of a k-nearest neighbors (k-NN) algorithm-based book recommendation system. The system was able to provide consumers with personalized book suggestions after being trained using a collection of user ratings and book information. The system's evaluation revealed that its suggestions had a high degree of precision, recall, and overall correctness. The algorithm proved successful in matching books with user's interests, as evidenced by the favorable user comments on the suggested books. To produce its suggestions, the algorithm also considered other data, such as genre, author, and book popularity. The book recommendation system, in general, was successful in offering users individualized and pertinent book selections.

```
Recommendations for Hurricane Bay:

1: 16 Lighthouse Road, with distance of 0.7278809052310367:
2: 204 Rosewood Lane, with distance of 0.7777173678117257:
3: Purity in Death, with distance of 0.8581887395483726:
4: Dark Water (Mira Romantic Suspense), with distance of 0.8881423745774621:
5: Haunted, with distance of 0.894654919771918:
```

In [ ]:

Figure 18: Achieved Result.

12   Conclusion

Coursework 2 is an assignment given to students to help them develop research skills and deepen their understanding of the topics listed above. The specific details of the assignment, such as the length and format, will vary depending on the instructor and the course in which it is being completed.

The topics listed in the assignment are all related to the field of artificial intelligence (AI), and students will be expected to conduct research and analysis on one or more of these topics in order to complete the assignment.

Problem solving and heuristic search are techniques used by AI systems to solve complex problems by using a combination of algorithms, data, and human-like reasoning. Adversarial search and games involve using AI to model strategic decision-making in situations where multiple agents are competing or collaborating to achieve their goals. Natural language processing is a field of AI that focuses on developing systems that can understand, interpret, and generate human language.

Machine learning is a subfield of AI that focuses on training algorithms to learn from data and make predictions or decisions without being explicitly programmed. There are two main types of machine learning: supervised learning and unsupervised learning. Supervised learning involves training algorithms on labeled data, where the correct output is provided, in order to make predictions on new, unseen data. Unsupervised learning involves training algorithms on unlabeled data and allowing the algorithms to find patterns and relationships within the data.

Recommendation systems are a type of AI-powered technology that uses machine learning algorithms to make personalized recommendations to users based on their past behaviors and preferences. These systems are commonly used in applications such as ecommerce, social media, and entertainment platforms to provide users with tailored content and suggestions.

In summary, Coursework 1 is an assignment that challenges students to conduct research and analysis on one or more topics within the field of AI. The specific details of the assignment will vary depending on the instructor and the course, but it will likely involve researching and writing about one or more of the topics listed above, such as problem solving, natural language processing, and recommendation systems. Working with artificial intelligence for the first time was challenging and exciting.

I have no prior knowledge of this topic. I have therefore conducted a great deal of research for my education and coursework. With the aid of collaborative filtering K-Means, I have selected a game recommendation system. I did a lot of research, and I included the most important findings in the report and section of the coursework. When the system needs a good and appropriate dataset, it will recommend games to the user. The complete report's research includes a type of recommendation system that describes how the system functions in each field. Similar system reviews are conducted in accordance with the subject matter I have chosen, examined, and reviewed, and where conclusions are drawn on the K-Mean algorithm.

12.1 Analysis of work done
Progress table

S.N	Task	Status
1	Research on Artificial intelligence Completed	Completed
2	Research on Recommendation System	Completed
3	Research on chosen topic (Book recommendation)	Completed
4	Problem statements of chosen topic	Completed
5	Research on work done method	Completed
6	Similar system review Completed	Completed
7	Review and analyzing existing work	Completed
8	Research solution of selected	Completed
9	Flowchart and pseudocode	Completed
10	Implementing coding	Completed
11	Testing project	Completed

Table 1: Table of Progress

12.2 How this solution address real world problem?

The majority of the software and programs we regularly use have recommendation systems built in. Utilizing a suggestion system might be quite beneficial to a user. This algorithm for recommendations can be used in online bookstores like ebook.

Libery can utilize this advise to guide them in selecting a book that they are most likely to enjoy rather than reading a book on their own without any recommendations. Time savings, better system confidence, and an increase in app revenue all benefit the user.

 -*

12.3 Further work

To our knowledge, the planning studies and the information we gather form the basis for this project.

After the research is finished, the entire report—including the analysis we made on the subject is implemented.

In the subsequent stage, I'll focus on putting this concept into action and documenting how it works.

I'll look for data sets and use every effort to identify the best ones and keep them rationally preserved. The document part, which includes the description of the solution and the AI generated pseudocode employed throughout, must be completed before moving on to the next section.

13 Bibliography

(n.d.). Retrieved from https://pypi.org/project/matplotlib/

.ijesrt.com. (2022). Retrieved from .ijesrt.com: http://www.ijesrt.com/issues%20pdf%20file/Archive-2017/November-2017/17.pdf

goodwillpublication.com. (2022). Retrieved from goodwillpublication.com: https://goodwillpublication.com/product/general-knowledge-1/

pritamaich. (2022). Retrieved from pritamaich: https://pritamaich.github.io/Books-Recommendation-System/

sajhakitab.com. (2022). Retrieved from sajhakitab.com: https://store.sajhakitab.com/product/the-intelligent-investor/

techtarget. (2023). Retrieved from techtarget: https://www.techtarget.com/whatis/definition/command-prompt

towardsdatascience. (2023). Retrieved from towardsdatascience.: https://towardsdatascience.com/pandas-i-read-csv-head-tail-info-and-describe-43b9b2736490

users.encs.com. (2022, 11 27). Retrieved from users.encs: https://users.encs.concordia.ca/~paquet/wiki/images/3/31/RUP.pdf

www.analyticsvidhya.com. (2022, 12 11). Retrieved from www.analyticsvidhya.com: https://www.analyticsvidhya.com/blog/2021/07/recommendation-system-understanding-the-basic-concepts/

www.computerhope.com/. (2023). Retrieved from www.computerhope.com/: https://www.computerhope.com/jargon/j/jupyter.htm

www.techopedia.com. (2023). Retrieved from www.techopedia.com: https://www.techopedia.com/definition/3533/python