BACS2003 ARTIFICIAL INTELLIGENCE

202301 Session, Year 2022/23

Assignment Documentation

Full Name: Tan Cherng Ming	
Student ID: 21WMR02971	
Programme: RSW2S3	
Tutorial Class: Group 2	
Project Title: Product Recommender System	
Module In-Charged: Collaborative Filtering Recommender System	
Lecturer: Dr. Goh Ching Pang	**Tutor: Dr. Ho Chuk Fong**

Other team members' data

No	Student Name	Module In Charge
1	Chai Wei Qi	Simple Recommender System
2	Oh Boon Suen	Content-Based Filtering Recommender System

1. Introduction

1.1. Problem Background

Product recommendation systems are created to offer customers tailored product recommendations based on their purchasing habits and interests. The user of the MyElectronic platform reported that they were having trouble finding the next product to purchase. Due to the similarity of all the products from the main interface, they were unable to choose which one to purchase. Additionally, respondents criticized their inability to predict which things will appeal to their tastes before making purchases. They lose a lot of time trying to find the desired product as a result of this issue. Users also mentioned that they occasionally had a poor experience purchasing items that they happened to randomly find on the online store. They are more likely to stop using the product after utilizing it for a while.

These issues must be resolved so that customers can receive recommended products that match their preferences. Every time a product ends, users wouldn't have to choose products by hand. In order to address the user's concerns, I have suggested adding a collaborative filtering and recommendation system to our system. The users can receive personalized material based on their ratings of other products by putting this recommender system into the MyElectronic platform. In addition, it can assist customers in selecting products to purchase by suggesting a list of products they would find interesting. As a result of people being satisfied with the things they are using, product sales may also rise. Due to their positive experiences, users may also suggest our platform to their friends and family, which could lead to a growth in user numbers. Additionally, it can improve the user's experience by displaying the platform's most popular product categories.

1.2. Objectives/Aims

The objective of our system is to attract more users to buy and use the products from the MyElectronic platform. Users may quickly find their preferred products by utilizing our recommender system. This might leave a positive image on our platform because people frequently return anytime they need to buy or use a product because they are happy with the things they have used. Additionally, this recommender system's goal is to give users a positive user experience. Users would like to utilize the recommended products since they matched their tastes if the recommender system were implemented. If a user had a positive experience, they would submit positive ratings or comments on our site and tell their friends and family about it. This may help our platform gain popularity and distinguish itself from rivals.

Besides, one of the goals is to shorten the user's search time for things to buy or utilize. The user can now utilize the recommendation area to find the next product to use thanks to the implementation of the recommender system. This offers the user ease and saves them time while looking for a product to utilize. The final goal of this recommender system is to give each user access to material that is specifically tailored to them. Based on the user's prior ratings of the products, the system proposes the user's preferred products. Future product preferences can be forecasted using algorithms, and this information is then shown on the platform.

1.3. Motivation

Product recommendation systems using machine learning can have significant commercialization value and social impacts. By providing more accurate recommendations that are more likely to be relevant to the user, product recommendation systems can increase user satisfaction and engagement with the platform. This can lead to increased user retention and loyalty, which can be valuable for businesses. Additionally, product recommendation systems can help users discover new products more easily and quickly, which can have positive social impacts by promoting product diversity and cultural exchange.

Machine learning is used in the industry in a variety of ways. In addition to product recommendation systems, machine learning is used for product generation. By analyzing user data and preferences, machine learning algorithms can generate personalized products that are tailored to the user's tastes. This can help users discover new products more easily and quickly, while also increasing engagement with the platform.

1.4. Timeline/Milestone

No.	Task	Date
1	Start the report by writing the Introduction section.	27/3/2023
2	Self-study of the chosen algorithms	29/3/2023
3	Identify the appropriate dataset for the product recommendation system together with the rest of the team.	1/4/2023
4	Load the data and display the notebook's fundamental data with the teammate through Google Meet.	6/4/2023
5	Perform data preprocessing together with other team members ● Drop any extra data/columns ● Analysing Exploratory Data	10/4/2023
6	Begin programming with Collaborative Filtering techniques.	16/4/2023
7	Submit the module to the group leader to compile all modules into a single Jupyter notebook.	29/4/2023
8	Finish the report.	2/4/2023
9	Submit the individual reports to the group leader	3/5/2023
10	Group Leader gathers all supporting materials and prototypes and submits them to Google Classroom	5/5/2023

2. Research Background

2.1. Background of the applications

A recommender system is a system that suggests items to users based on assumptions about their future preferences. Users are given the most pertinent content recommendations in order to maximize the likelihood that they would buy the product. Product recommender systems are intended to do so based on their purchasing patterns and preferences. These systems analyze user data using machine learning algorithms to produce recommendations that are more likely to be pertinent to the user. Product recommendation systems can be divided into three categories: simple-based, content-based, and collaborative.

The concept of recommender systems was first mentioned in a technical report as a “digital bookshelf” in 1990 by Jussi Karlgren at Columbia University (Caulfield, n.d.). Then, starting in 1994, Jussi Karlgren, then at SICS, and research teams led by Pattie Maes at MIT, Will Hill at Bellcore, and Paul Resnick, also at MIT, who's work with GroupLens won the 2010 ACM Software Systems Award, implemented it at scale and worked through it in technical reports and publications. Information retrieval and cognitive science studies laid the foundation for recommender systems. Its first manifestation was the Usenet communication system created by Duke University in the second half of the 1970s, where users were able to share textual content with each other (Apáthy, 2021).

The popularity of product recommendation systems has grown over the past few years as a result of the expansion of digital platforms and the accessibility of vast volumes of user data. By providing more accurate recommendations that are more likely to be relevant to the user, product recommender systems can increase user satisfaction and engagement with the platform. This can lead to increased user retention and loyalty, which can be valuable for businesses.

There are many companies that use recommender systems to improve customer experience and increase their sales revenue. Some examples include Amazon, Netflix, Walmart, and eBay (H, 2022). These companies collect and analyze demographic data from customers and add it to information from previous purchases, product ratings, and user behavior. They then use machine learning algorithms to generate personalized recommendations that are more likely to be relevant to the user. By providing more accurate recommendations that are more likely to be relevant to the user, product recommender systems can increase user satisfaction and engagement with the platform. This can lead to increased user retention and loyalty, which can be valuable for businesses (Underwood, 2019).

2.2. Analysis of selected tool with any other relevant tools

Tools comparison	Remark	Jupyter Notebook	Java SE	Spyder
Type of license and open source license	State all types of license	BSD License	Oracle Technology Network License	MIT License
Year founded	When is this tool being introduced?	2014	1995	2009
Founding company	Owner	Jupyter	Sun Microsystem	Spyder Project Contributor
License Pricing	Compare the prices if the license is used for development and business/commercialization	No additional license pricing	No additional license pricing	No additional license pricing
Supported features	What features that it offers?	Supports over 100 programming languages and has a variety of features such as interactive widgets, data visualization tools, and support for LaTeX equations.	Object-oriented programming, multithreading, networking, and security features.	Supports scientific computing libraries such as NumPy and SciPy, and has a variety of features such as code completion, debugging tools, and support for multiple languages.
Common applications	In what areas this tool is usually used?	Data analysis, machine learning, scientific computing, and data visualization.	Desktop applications, web applications, mobile applications, and games.	Data analysis, scientific computing, machine learning, and data visualization
Customer support	How the customer support is given, e.g. proprietary, online community, etc.	Online Community	Online Blog	Online customer service
Limitations	The drawbacks of the software	Code-style correction is not provided and only runs cells one by one	Difficult to use for beginners and doesn't support the development of complex UI	Slow when working with large datasets and can be difficult to use for beginners

2.3. Justify why the selected tool is suitable

Jupyter Notebook is a fantastic tool for developing the system As an open-source web application that enables the creation and sharing of documents with live code, equations, visualizations, and narrative text. It supports over 100 programming languages and has a variety of features such as interactive widgets, data visualization tools, and support for LaTeX equations. Jupyter Notebook is commonly used for data analysis, machine learning, scientific computing, and data visualization. It can display data visualization, which is appropriate for my project because it calls for displaying data visualization to analyze the data we utilized before running algorithms, such as bar charts, graphs, and pie charts. It is also great for developing systems because it allows users to easily test and debug code in real time. It works well for my project because I can code faster and find issues much more rapidly.

3. Methodology

3.1. Description of dataset

The dataset used for this project is the **review** and **product metadata** for the **electronics** category adapted from Amazon Review Data 2018 (Jianmo, 2019; Jianmo, n.d.). The dataset could be also accessed through Kaggle (Saurav, 2022) or Recommender Systems and Personalization Datasets (Julian, n.d.).

The dataset consists of two files which are user_ratings.csv and electronic_products.json.
For my module, I have used both of the files.

user_ratings.csv

- Consists of the user ratings for the electronic products.
- Data dictionary:

Variables	Data Types	Description
userId	string	The unique identifier that represents the user
productId	string	The unique identifier that represents the product
Rating	float	The rating for the product given by the user
timestamp	integer	Unix time that represents the date and time but in the computer's clock format

electronic_products.json

- Consists of the metadata of electronic products.
- Data dictionary:

Variables	Data Types	Description
category	string	List of categories that the product belongs to
tech1	string	The first technical detail table of the product
description	string	The description of the product
fit	string	-
title	string	The name of the product
also_buy	string	A product list that indicates the user who bought the product also bought the list of product
tech2	string	The second technical detail table of the product

brand	string	The brand name for the product
feature	string	Bullet-point format features of the product
rank	string	Information on sales rank that shows how the product is doing in comparison to other products in the same category
also_view	string	A product list that indicates the user who viewed the product also viewed the list of product
main_cat	string	The main category assigned for the product
similar_item	string	HTML code that shows the similar product
date	string	The date that the product is added
price	string	The price of the product
asin	string	The unique identifier that represents the product
imageURL	string	The URL of the product image
imageURLHighRes	string	The URL of the product image with a higher resolution
details	string	The relevant details for the product

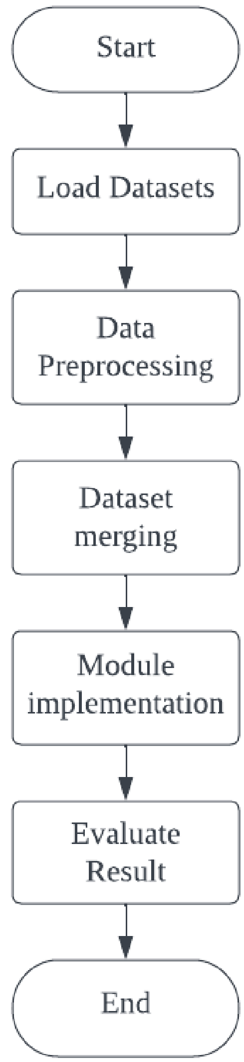
3.2. Applications of the algorithm(s)

Collaborative filtering technique is being chosen for this product recommender system. It is a kind of recommender system that makes suggestions for products depending on the tastes of other users. In other words, collaborative filtering systems recommend items that users with similar tastes have liked in the past.

Collaborative filtering is used for product recommender systems because it is based on the idea that if two users have similar preferences for products, they are likely to enjoy similar products. It works by analyzing the preferences of multiple users and identifying patterns in their preferences. For example, if two users have both liked the same products in the past, the system will recommend other products that one user has liked to the other user.

There are two main types of collaborative filtering: **user-based** and **item-based**. User-based collaborative filtering suggests products based on those of like users. Item-based collaborative filtering recommends items based on their similarity to items the user has already liked.

3.3. System flowchart/activity diagram



3.4. Proposed test plan/hypothesis

First, two datasets from Kaggle and the Recommender Systems and Personalization Datasets have been modified. To reduce data redundancy and boost the accuracy of the recommender system, we performed data preprocessing by dividing the file into a few smaller files, removing unused columns, decreasing the amount of data in the dataset, and deleting duplicate data. Since this module is using Collaborative Filtering, I used **`pandas.pivot_table`** to build a pivot table in the style of a spreadsheet as a data frame. For the item-based filtering, I utilized **`pandas.DataFrame.corrwith`** function which calculates the pairwise correlation between products. The correlation-based product is then shown in a data frame descendingly. The ratings are then added to the data frame using the **`join()`** method. Lastly, 20 products with more than 50 ratings and a better correlation value will be selected for recommendation.

For the user-based filtering, I used the **numpy.transpose()** function to construct the matrix transpose by flipping the array's axes. the **pandas.DataFrame.corrwith** function is again used to calculate the pairwise correlation between the users based on the product ratings. **pandas.DataFrame.iloc** function is also being used to select a specific row or column from the data set. The user id of the user who is the most similar to another user is also extracted using the **index.values.tolist()** function. In the final step, data slicing, products with ratings more than 3.0 will be chosen to be suggested.

4. Result

4.1. Results

Item-based filtering

1. Find similar products based on one product. **dataframe.corrwith()** is used to calculate the pairwise correlation between products

```
In [18]: #One item is selected
         users_ratings = user_product_matrix['Koss Porta Pro On Ear Headphones with Case, Black / Silver']
         users_ratings.head(10)

Out[18]: userId
         A001944026UMZ8T3K5QH1    0.0
         A00766851QZZUBOVF4JFT    0.0
         A01255851ZO1U93P8RKGE    0.0
         A014623426J5CM7M12MBW    0.0
         A01580702BRW77PSJ9X34    0.0
         A01852072Z7B68UHLI5UG    0.0
         A0266076X6KPZ6CCHGVS     0.0
         A0293130VTX2ZXA70JQS     0.0
         A030530627MK66BD8V4LN    0.0
         A03279253KKB83JP34CU     0.0
         Name: Koss Porta Pro On Ear Headphones with Case, Black / Silver, dtype: float64

In [19]: #Calculate the correlation
         similar_product = user_product_matrix.corrwith(users_ratings)
         similar_product

Out[19]: title
         "The Arp Atlas of Peculiar Galaxies", Hardcover Book by Kanipe/Webb
         -0.000618
         //
         -0.000618
         1 X Professional Ultra SanDisk MicroSDXC 32GB (32 Gigabyte) Card for GoPro HERO3 Upgrade is custom formatted and rated for high
         speed, lossless recording!. (XD UHS-I Class 10 Certified 30MB/sec+)    -0.000874
         10 Ivory Leviton 1-Gang Decora GFI GFCI Cover Thermoset Wallplates 80401-I
         -0.000618
         100" Portable Pull Up 4:3 Floor Screen Aluminium Structure Projector Projection
         -0.000618

         ...
         iPod Touch 5th Generation Silicone Skin Case Black
         -0.001047
         iREZ KritterUSB Webcam PC/Mac (Blue)
         -0.000874
         jWIN JTP10 Caller ID Box
         -0.000750
         my-Vox Digital Voice Recorder Springboard Module (Graphite)
         -0.001322
         nik Color Efex Pro 2.0 Standard Edition
         -0.000848
         Length: 5457, dtype: float64
```

2. Convert the result into a data frame and display it in descending order.

```
In [21]: #Sort the product with correlation descendingly
         similar_product.sort_values(by = 'Correlation', ascending = False).head(10)
```

Out[21]:

title	Correlation
Koss Porta Pro On Ear Headphones with Case, Black / Silver	1.000000
Premium 6 Foot High Speed HDMI Cable for your Sharp Aquos LED-LCD HDTV ! Supports: 1080p-2160p, 4K, 3D, Deep Color, TrueHD, CL3, and 800Hz technologies.	0.023813
Koss SportaPro Stereo Headphones	0.021189
Pine Technology 2SM200C+ D'Music Portable MP3-CD Player	0.016303
Imation SuperDisk 120MB Parallel Port External Drive	0.012771
Panasonic SLS362C Portable CD Player Car Kit with Remote	0.012121
Minolta Freedom Zoom Supreme EX 35mm Camera	0.012114
Barbie Polaroid I-Zone Instant Pocket Camera	0.011321
Brother 3/4 Inch x 26.2 Feet Black on White Tape with Super Strong Adhesive (TZS241)	0.008498
Diamond SupraExpress USB (PC/Mac)	0.008415

3. Count the number of ratings for the products and join the result to the previous data frame.

```
In [22]: #Count number of rating for the title
         df_rating = pd.DataFrame(product_ratings.groupby('title')['Rating'].count())

In [23]: recommend_product = similar_product.join(df_rating['Rating']).sort_values(by = 'Correlation', ascending = False)
         recommend_product
```

Out[23]:

title	Correlation	Rating
Koss Porta Pro On Ear Headphones with Case, Black / Silver	1.000000	2075
Premium 6 Foot High Speed HDMI Cable for your Sharp Aquos LED-LCD HDTV ! Supports: 1080p-2160p, 4K, 3D, Deep Color, TrueHD, CL3, and 800Hz technologies.	0.023813	1
Koss SportaPro Stereo Headphones	0.021189	209
Pine Technology 2SM200C+ D'Music Portable MP3-CD Player	0.016303	4
Imation SuperDisk 120MB Parallel Port External Drive	0.012771	3
...	...	...
Panasonic Headphones On-Ear Lightweight with XBS RP-HT21 (Black & Silver)	-0.017951	1692
Cisco-Linksys BEFSR41 EtherFast Cable/DSL Router with 4-Port 10/100 Switch	-0.018448	1006
Bushnell Powerview Compact Folding Roof Prism Binocular	-0.018641	1030
Sony MDRV6 Studio Monitor Headphones with CCAW Voice Coil	-0.019245	1586
VideoSecu 24" Long Arm TV Wall Mount Low Profile Articulating Full Motion Cantilever Swing Tilt wall bracket for most 22" to 55" LED LCD TV Monitor Flat Screen VESA 200x200 400x400 up to 600x400mm MAH	-0.019755	1051

5457 rows × 2 columns

4. Remove the products with a rating count lower than or equal to 50 and keep the first 20 products on a list.

```
In [24]: # Recommend top 20 product that has > 50 ratings
recommend_product = recommend_product[recommend_product['Rating'] > 50].sort_values(by = 'Correlation', ascending = False)
recommend_product.head(20)
```

Out[24]:

title	Correlation	Rating
Koss Porta Pro On Ear Headphones with Case, Black / Silver	1.000000	2075
Koss SportaPro Stereo Headphones	0.021189	209
Belkin USB 4-Port Hub with 6FT Cable/Power Supply	0.002462	56
Koss VC20 Volume Control	0.000489	430
Sennheiser HD 600 Open Back Professional Headphone	-0.000527	136
Viking 64 MB SmartMedia Card (SSFDC3/64)	-0.000863	51
Koss QZ-99 Noise Reduction Stereophone	-0.001926	121
Viking InterWorks IntelliFlash - Card reader (SM, PC Card, CF) - USB	-0.002116	88
RCA RC5220P DVD Player	-0.002229	53
Iomega 31310 Zip 250 MB USB-Powered Drive	-0.002368	95
NETGEAR 8-Port Fast Ethernet Unmanaged Switch, Desktop, ProSAFE Lifetime Protection (FS108NA)	-0.002714	74
Belkin A3L791-S 50-Foot RJ45 CAT 5e Snagless Molded Patch Cable (Gray)	-0.002884	89
KB Gear JamCam 0.3MP Digital Camera, Silver	-0.002923	94
Sony MDR-V300 Monitor Series Headphones with Folding Design	-0.003171	94
Allsop CD Laser-Lens Cleaner	-0.003233	109
Belkin F3U133-06 Pro Series Hi-Speed USB Cable (Six-Feet)	-0.003351	128
Koss Pro-4AA Studio Quality Headphones	-0.003361	162
D-Link DSB-H4 4-Port USB 1.1 Hub	-0.003522	64
Sony MVC-FD73 0.3MP Mavica Digital Camera w/ 10x Optical Zoom	-0.003632	70
Midland 18-258 40-Channel Glass-Mount CB Antenna	-0.003885	51

```
In [25]: #Extract 20 product and make the recommended items a list
recommend_product = recommend_product.iloc[1:21]
products = recommend_product.index.values.tolist()
products
```

```
Out[25]: ['Koss SportaPro Stereo Headphones',
 'Belkin USB 4-Port Hub with 6FT Cable/Power Supply',
 'Koss VC20 Volume Control',
 'Sennheiser HD 600 Open Back Professional Headphone',
 'Viking 64 MB SmartMedia Card (SSFDC3/64)',
 'Koss QZ-99 Noise Reduction Stereophone',
 'Viking InterWorks IntelliFlash - Card reader ( SM, PC Card, CF ) - USB',
 'RCA RC5220P DVD Player',
 'Iomega 31310 Zip 250 MB USB-Powered Drive',
 'NETGEAR 8-Port Fast Ethernet Unmanaged Switch, Desktop, ProSAFE Lifetime Protection (FS108NA)',
 'Belkin A3L791-S 50-Foot RJ45 CAT 5e Snagless Molded Patch Cable (Gray)',
 'KB Gear JamCam 0.3MP Digital Camera, Silver',
 'Sony MDR-V300 Monitor Series Headphones with Folding Design',
 'Allsop CD Laser-Lens Cleaner',
 'Belkin F3U133-06 Pro Series Hi-Speed USB Cable (Six-Feet)',
 'Koss Pro-4AA Studio Quality Headphones',
 'D-Link DSB-H4 4-Port USB 1.1 Hub',
 'Sony MVC-FD73 0.3MP Mavica Digital Camera w/ 10x Optical Zoom',
 'Midland 18-258 40-Channel Glass-Mount CB Antenna',
 'Toshiba SD2700 DVD Player']
```

User-based filtering

1. Use **`numpy.transpose()`** to construct the matrix transpose by flipping the array's axes.

```
In [26]: #Transpose the pivot table
         product_user_matrix = user_product_matrix.transpose()
         product_user_matrix.head()
```

Out[26]:

userId	A001944026UMZ8T3K5QH1	A00766851QZZUBOVF4JFT	A01255851ZO1U93P8RKGE	A014623426J5CM7M12MBW	A01580702BRW77PSJ9X34	A0185207
title						
"The Arp Atlas of Peculiar Galaxies", Hardcover Book by Kanipe/Webb	0.0	0.0	0.0	0.0	0.0	
//	0.0	0.0	0.0	0.0	0.0	
1 X Professional Ultra SanDisk MicroSDXC 32GB (32 Gigabyte) Card for GoPro HERO3 Upgrade is custom formatted and rated for high speed, lossless recording!. (XD UHS-I Class 10 Certified 30MB/sec+)	0.0	0.0	0.0	0.0	0.0	
10 Ivory Leviton 1-Gang Decora GFI GFCI Cover Thermoset Wallplates 80401-I	0.0	0.0	0.0	0.0	0.0	
100" Portable Pull Up 4:3 Floor Screen Aluminium Structure Projector Projection	0.0	0.0	0.0	0.0	0.0	

5 rows × 73071 columns

2. Find similar users based on one user. **`dataframe.corrwith()`** is used to calculate the pairwise correlation between products. Convert the result into a data frame

```
In [27]: # One user is selected
         user_title_ratings = product_user_matrix['A231WM2Z2JL0U3']
         user_title_ratings.head(5)

Out[27]: title
         "The Arp Atlas of Peculiar Galaxies", Hardcover Book by Kanipe/Webb
         0.0
         //
         0.0
         1 X Professional Ultra SanDisk MicroSDXC 32GB (32 Gigabyte) Card for GoPro HERO3 Upgrade is custom formatted and rated for high
         speed, lossless recording!. (XD UHS-I Class 10 Certified 30MB/sec+)    0.0
         10 Ivory Leviton 1-Gang Decora GFI GFCI Cover Thermoset Wallplates 80401-I
         0.0
         100" Portable Pull Up 4:3 Floor Screen Aluminium Structure Projector Projection
         0.0
         Name: A231WM2Z2JL0U3, dtype: float64

In [28]: #Calculate the correlation
         similar_users = product_user_matrix.corrwith(user_title_ratings)

         # Create a dataframe
         similar_users = pd.DataFrame(similar_users, columns = ['Correlation'])
         similar_users.head(10)

Out[28]:
```

userId	Correlation
A001944026UMZ8T3K5QH1	-0.001647
A00766851QZZUBOVF4JFT	-0.001647
A01255851ZO1U93P8RKGE	-0.001647
A014623426J5CM7M12MBW	-0.001647
A01580702BRW77PSJ9X34	-0.001647
A01852072Z7B68UHLI5UG	-0.001647
A0266076X6KPZ6CCHGVS	-0.001647
A0293130VTX2ZXA70JQS	-0.001647
A030530627MK66BD8V4LN	-0.001647
A03279253KKB83JP34CU	-0.001647

3. Display the data frame in descending order. **`Dataframe.iloc[]`** is used to select the first 20 users with the highest correlation and exclude the selected user.

```
In [29]: #Sort the user with correlation descendingly
         most_similar_users = similar_users.sort_values(by = 'Correlation', ascending = False).iloc[1:21]
         most_similar_users
```

Out[29]:

userId	Correlation
A2MAXHANV8EETT	0.238092
A33H0WC9MI8OVW	0.207888
A105GWGM7PDAI2	0.199007
A3L60UJZQ0O0R5	0.194024
A7IKOAH6M8WAC	0.184025
A2WHH8X74OY0YS	0.174245
A3K2LLDVHSP7P1	0.174245
AUI4PY52AUM77	0.174245
A3OCHH22N0W1AR	0.174245
AKYSM32DSG2TK	0.174245
A34N0OYW8076L6	0.174245
A4Z2GNX8NDP4A	0.174245
A1MZ2BJHUSTW0W	0.174245
A1RVSPTDB45J2M	0.174245
AZQ8HGO3Z4T	0.174245
AYW1O00QM271D	0.174245
A2SEEDV7NNPIBG	0.174245
ABH7Q75AT8R21	0.173179
A1H8F4OVEM09DO	0.173179
A1XFVNFGS78TX1	0.173179

4. **`index.values.tolist()`** is used to obtain the user ID of the user who, according to the correlation score, is the most comparable to the target user and has the highest correlation with the target user. Display the product that the user rated.

```
In [30]: #Extract the first most similar user
         user_list = most_similar_users.index.values.tolist()
         user_list[0]

Out[30]: 'A2MAXHANV8EETT'

In [31]: #Product that are rated the user
         recommendation = product_ratings[product_ratings['userId'] == user_list[0]]
         recommendation
```

Out[31]:

	userId	productId	Rating	title	main_cat
44901	A2MAXHANV8EETT	B000023VUL	5.0	D-Link DSB-H4 4-Port USB 1.1 Hub	All Electronics
48259	A2MAXHANV8EETT	B000030067	2.0	Cisco-Linksys EPSX3 EtherFast 10/100 3-port Pr...	All Electronics
61107	A2MAXHANV8EETT	B00004SB92	5.0	Cisco-Linksys BEFSR41 EtherFast Cable/DSL Rout...	All Electronics
73299	A2MAXHANV8EETT	B00004TEN2	3.0	Viking InterWorks IntelliFlash - Card reader (...	Computers
95841	A2MAXHANV8EETT	B00004XRDB	2.0	Cisco-Linksys EFSP42 EtherFast 10/100 2-Port S...	All Electronics
137381	A2MAXHANV8EETT	B00005ARK3	3.0	Cisco-Linksys BEFW11S4 Wireless-B Cable/DSL Ro...	All Electronics
143731	A2MAXHANV8EETT	B00005AW1H	4.0	Linksys WUSB11 Wireless-B USB Network Adapter v4	All Electronics

5. Remove the products with a rating lower than or equal to 3.0 and display the remaining products on a list.

In [32]:
```
#DataFrame slicing : product with the rating > 3.0
recommendation = product_ratings.loc[(product_ratings['userId'] == user_list[0]) &
                                     (product_ratings['Rating'] > 3),
                                     ['title', 'Rating']]
recommendation
```

Out[32]:

	title	Rating
44901	D-Link DSB-H4 4-Port USB 1.1 Hub	5.0
61107	Cisco-Linksys BEFSR41 EtherFast Cable/DSL Rout...	5.0
143731	Linksys WUSB11 Wireless-B USB Network Adapter v4	4.0

In [33]:
```
recommendation = recommendation.set_index('title')
recommendation_list = recommendation.index.values.tolist()
print('List to recommend')
recommendation_list
```

```
List to recommend
```

Out[33]:
```
['D-Link DSB-H4 4-Port USB 1.1 Hub',
 'Cisco-Linksys BEFSR41 EtherFast Cable/DSL Router with 4-Port 10/100 Switch',
 'Linksys WUSB11 Wireless-B USB Network Adapter v4']
```

4.2. Discussion/Interpretation

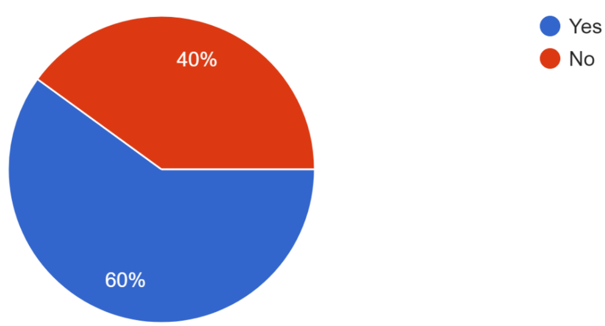
Collaborative Filtering is able to learn to provide better recommendations as more information about users is collected. The algorithms are able to execute and produce the desired output whether it is user-based or item-based. For both methods of recommendation, the algorithms are able to recommend the products with pairwise correlation values. However, the recommended product didn't have a high correlation value for both approaches.

In order to assess the system's effectiveness, we published a Google form and gathered 10 respondents to vote on the appropriateness of the result.

<u>Item-based:</u>

Is the system recommending the right items?   Assume that you have purchased this product: Koss Porta Pro On Ear Headphones with Case   This is wh... Tilt wall bracket for most 22" to 55" LED LCD TV

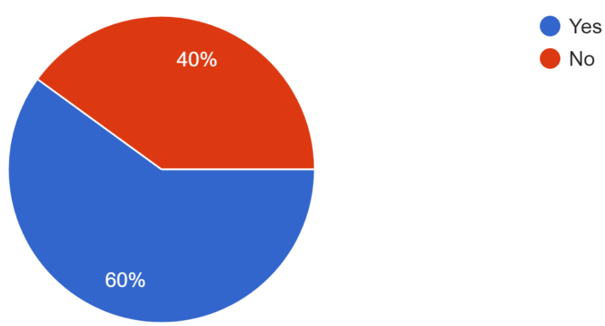
10 responses



<u>User-based:</u>

Is the system recommending the right items?   Assume that you have purchased this product: - Belkin F2A046-06 IEEE 1284 PC A-B P...inksys WUSB11 Wireless-B USB Network Adapter v4

10 responses



From the pie charts above, we can see that 60% of people agree that the recommended products based on item-based filtering are relevant. The same goes for user-based, 60% of people agree that the products that are recommended using user-based filtering are relevant. In conclusion, the system is recommending products that are partially agreed upon by the public.

5. Discussion and Conclusion

5.1. Achievements

The project's purpose was accomplished in this module. The MyElectronic platform benefits greatly from the collaborative filtering recommender system. By recommending their preferred products, it makes it easier for consumers to find products to buy.

This project has achieved all of its goals because it will increase user satisfaction and the recommender system can also improve the user experience. With item-based filtering, the customer is shown a list of products based on a certain item as well as the top items they should purchase based on their favorite product list. The user now has a variety of options when selecting things to buy.

In addition, this project achieves the goal by giving each user access to personalized material. The suggested products are determined by the actions of the particular user. By doing this, it is ensured that the items displayed in the primary interface of the online shopping platform are based on their individual tastes.

Additionally, since the user spends less time looking for things to buy or use, objectives have also been achieved. The user will see a list of suggested products so they can quickly find items to use without having to look for them.

Furthermore, users are spending greater time on our online purchasing platform. The goods that are suggested are based on user ratings, therefore there is a high likelihood that the target user will be interested in utilizing them. This implies that they will use the merchandise extensively. They would probably return to this platform to purchase or use things.

5.2. Limitations and Future Works

The dataset used in the project is also limited to 100,000 products. While this is a large dataset, it may not represent all products. This means the recommendations may not be as accurate as they could be if a larger and more diverse dataset was used. Thus, the project could be expanded to include a larger and more diverse dataset of products in the future. As the result shown in 4.2, the number of people who agreed with the result is only slightly higher. This could be due to insufficient data used or a lack of diversity in the data. Hence, by increasing the number of the dataset, it would help to ensure that the recommendations are more accurate and persuasive.

Besides, the project also didn’t consider all product features such as user’s browsing history, user’s purchase history, and demographic information. These features can be important in determining whether a user will likely purchase a particular product. By not taking these features into account, the recommendations may not be as accurate as they could be. For future improvement, the project could be expanded to include additional features such as user’s browsing history, user's purchase history, and demographic information to ensure that the recommendations are more accurate and take into account all aspects of the product.

Finally, the project doesn't take into account changing user preferences. Users' product preferences can change over time and the system doesn't have a way to adapt to these changes. This means that the recommendations may become less accurate over time if users' product preferences change. The project could be expanded to include machine learning algorithms that can adapt to changes in user preferences over time for future improvement. Hence, it ensures that the recommendations remain accurate even as users' product preference change.

Reference & Source

Source of Dataset

https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews

https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

Tool

Jupyter Notebook

References

Alsobhi, A. and Amare, N. (2022). Ontology-Based Relational Product Recommendation System. *Computational and Mathematical Methods in Medicine*, 2022, pp.1–11. doi:https://doi.org/10.1155/2022/1591044.

Apáthy, S. (2021). *History of recommender systems: overview of information filtering solutions*. [online] onespire.net. Available at: https://onespire.net/news/history-of-recommender-systems/.

Caulfield, B. (n.d.). *What's a Recommender System? – Carrushome.com*. [online] Carrushome. Available at: https://www.carrushome.com/en/whats-a-recommender-system/#:~:text=Jussi%20Karlgren%20formulated%20the%20idea%20of%20a%20recommender [Accessed 23 Apr. 2023].

H, N. (2022). *Best 10 AI Product Recommendation Systems of eCommerce*. [online] AI ML GEN. Available at: https://aimlgen.com/ai-product-recommendation-systems/ [Accessed 25 Apr. 2023].

Jianmo, N., Jiacheng, L. and Julian, M. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. Empirical Methods in Natural Language Processing (EMNLP), 2019, 188-197. doi: 10.18653/v1/D19-1018.

Jianmo, N. (n.d.). Amazon review data (2018). [online] cseweb.ucsd.edu. Available at: https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/ [Accessed 23 Apr. 2023].

Julian, M. (n.d.). Recommender Systems and Personalization Datasets. [online] cseweb.ucsd.edu. Available at: https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews [Accessed 23 Apr. 2023].

Saurav, A. (2022). Amazon Product Reviews. [online] kaggle. Available at: https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews [Accessed 23 Apr. 2023].

Underwood, C. (2019). *Use Cases of Recommendation Systems in Business – Current Applications and Methods | Emerj*. [online] Emerj. Available at: https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/ [Accessed 25 Apr. 2023].