TUNKU ABDUL RAHMAN UNIVERSITY COLLEGE

BEYOND EDUCATION

BACS2003 ARTIFICIAL INTELLIGENCE

202301 Session, Year 2022/23

Assignment Documentation

Full Name: Chai Wei Qi
Student ID: 21WMR02926
Programme: RSW2S3
Tutorial Class: Group 2
Project Title: Product Recommender System
Module In-Charged: Simple Recommender System

Other team members' data

No	Student Name	Module In Charge
1	Oh Boon Suen	Content-based Filtering Recommender System
2	Tan Cherng Ming	Collaborative Filtering Recommender System

Lecturer: Dr. Goh Ching Pang	**Tutor : Dr. Ho Chuk Fong**

1. Introduction

1.1. Problem Background

Recently, a new e-commerce business, **MyElectronic**, has launched their web-based online shopping platform. The website focuses on providing its customers with a huge selection of electronic goods. However, only a few months after the website's launch, reports have been showing a drop in sales performance and site visit rate, with the sales volume failing to meet the business' expected ROI (Return on Investment).

According to the website's users feedback, most of them find it troublesome and time-consuming to discover the products that match their individual needs and interests on MyElectronic's website. They find the vast array of products available overwhelming, and they often experience boredness or unsatisfactoriness in the process of navigating and browsing on the site. Specifically, customers have expressed concerns about the time-consuming and tedious task of scrolling through the product list in the product catalogue page to find the desired items. This has resulted in customer frustration and reduced retention, negatively affecting the overall shopping experience. To solve the problem, the business is dedicated to enhancing its website's usability and efficiency so that customers could discover trendy and popular products, or easily find the products they like.

A recommender system (RS) could effectively help the website to address the issue of information overload and enhance the product discovery process (Sharma et al., 2021). It is able to provide filtered information and personalised suggestions to the users, and would not only assist the users in finding information but also benefits business or content providers with growth of profit and sales. This technique has been widely adopted by innumerable applications, including e-commerce platforms, social media, video-sharing platforms, and many more (Jiawei et al., 2023). In this era of information explosion, a system to manage information, such as the RS, is crucial to tackle the issues of information overload.

Therefore, in this project, we will be developing a recommender system for MyElectronic. It will include 3 different approaches for different scenarios, such as simple, content-based filtering, and collaborative filtering. In this module, a simple product recommender system is proposed, which will recommend products to users (especially new users who are discovering products) based on the products' popularity (other users' ratings). Additionally, this simple recommender system will provide a filter feature to search the 'hot', 'trendy', or 'popular' products based on certain product name, category, brand, or price range entered by the users. By improving the product discovery experience, the business can better meet its customers' needs and increase its sales performance.

1.2. Objectives/Aims

Improve System Quality with Recommendation Engine

The long-term health and scalability of MyElectronic's web-based online shopping platform depend heavily on the system quality. The implementation of the recommendation system (RS) targets to enhance one of the metrics of system quality, efficiency, with its establishment of UI features that help users to achieve their task, which is to discover or find specific products that he or she is interested in, with less time and resources. Efficiency could be achieved by the RS as it provides users with a user-friendly interface to help them discover popular product suggestions or personalised product recommendations based on users' ratings and products' metadata. Therefore, RS could improve the overall system quality, which will bring additional value to customers and enhance MyElectronic's competitiveness in the market.

Increase Customer Lifetime Value (CLV)

Besides, the adoption of the recommendation system also aims to increase the customer lifetime value (CLV) which will effectively increase the number of profitable customers. A well-designed recommendation engine can increase customer lifetime value (CLV) by increasing customer loyalty, retention, and repeat purchases. By providing trending and personalised recommendations with the RS, MyElectronic can create a personalised shopping experience, which brings customer satisfaction and leads to improvement of average number of purchases or purchase frequency. These are important for an e-commerce business to increase the chances of customer loyalty and retention to achieve a higher CLV, as loyal customers tend to purchase more frequently and spend more.average number of purchases or purchases (*Raven360,* 2023).

Encourage Upselling and Cross-Selling

The third objective of applying the recommendation system is to encourage the upselling and cross-selling. The RS recommends related (cross-selling) or supplementary (upselling) electronic products based on the customer's purchases or interests (Sophia, 2022). For example, after a purchase of a certain product is made, a simple recommender could recommend other popular products associated with the purchased product's category, and a content-based or collaborative filtering approach could recommend products relevant to the purchased product. This in turn increases the average order value and average number of purchases of the website. Therefore, with a RS, the sales performance of MyElectronic could be increased to meet the business' expected ROI.

1.3. Motivation

In terms of potential commercialisation value, a RS could help businesses to understand user preferences and optimise their product inventory, leading to improved user satisfaction and increased sales. Not only could we help MyElectronics to stabilise their business growth, we could also sell the recommendation system as software service to other businesses which do online retail. If a business has the knowledge of which products have the high demands and high ratings, they could produce more similar products to increase the revenue. At the same time, those products with higher sales and ratings are more likely to have a higher quality, thus it would also be preferred by the customers and benefit them. Moreover, with the information provided by the RS, retail business could reduce the chances of out-of-stock products by adjusting the product inventory based on the RS's results. When a business continues to provide quality and attractive products, and has good stock management, they could also improve their reputation and raise their brand awareness. Simultaneously, customer satisfaction could also be increased. By all that, RS could successfully create a virtuous circle to the market, and retail businesses will be pleased to adopt the RS service.

In addition to potential commercialisation value, a RS can also have significant social impacts by facilitating users to find the right product effectively and efficiently. With that, a RS can reduce the time and effort required for users to do product searching and provide benefits to users who may have busy schedules or limited time to spend on shopping. Moreover, by recommending trendy products, the RS can help users discover new and exciting products that may enhance their life quality. This is because those popular products usually have a higher quality and could effectively solve users' pain points. As customers become more aware of quality requirements and their own demands, they are less likely to settle for lower quality products provided by businesses. Therefore, by providing high-quality product recommendations, the RS can encourage businesses to prioritise quality and customer satisfaction, leading to a more sustainable and responsible retail industry.

1.4. Timeline/Milestone

No	Task	Start Date	End Date
1	Determine the problem background and start the writing of the introduction section.	27/3/2023	28/3/2023
2	Study on recommender systems. Distribute modules among teammates.	28/3/2023	29/3/2023
3	Study simple recommender systems and write research background.	29/3/2023	1/4/2023
4	Discover and determine a suitable dataset for the product recommendation system with the team.	1/4/2023	2/4/2023
5	Load the dataset and display the fundamental data in Jupyter notebook with the team in Google Meet.	6/4/2023	6/4/2023
6	Perform data preprocessing together with the team. ● Remove duplicating records ● Drop extra columns (features) ● Explore and analyse data	10/4/2023	10/4/2023
7	Develop the prototype of the simple recommender system and write the Methodology section.	16/4/2023	25/4/2023
8	Prepare and distribute a Google Form for evaluation purposes.	26/4/2023	1/5/2023
9	Write the Result section.	1/5/2023	2/5/2023
10	Write the Discussion and Conclusion section.	2/5/2023	3/5/2023
11	Finalise the individual report.	3/5/2023	4/5/2023
12	Collect and finalise all documentations and prototypes of the team. Submit them to the Google Classroom.	5/5/2023	5/5/2023

2. Research Background

2.1. Background of the applications

A Recommender System (RS) takes in input about users' preferences and/or product information, analyses them using a certain algorithm, and outputs predictions about the items' ranking as a bundle (Milano, Taddeo and Floridi, 2020). In the mid 1970s, recommender systems had evolved as an independent research area in Duke University (Sharma and Singh, 2016). Later during the 1990s, the RS concept was introduced and was first alluded to by Jussi Karlgren at Columbia University in a specialised report as a "digital bookshelf" (Jussi, 2019). In 1992, the first recommender system which used collaborative filtering on a mail system came into existence at the Xerox Palo Alto Research Centre (Jannach et al., 2021).

There are several approaches to develop a recommender system, sometimes they are used together in a system for different scenarios. Some of the approaches include simple recommendation, content-based filtering, collaborative-based filtering, hybrid filtering, knowledge-based filtering, metadata-based filtering, demographic-based filtering, and many others (Roy, Sharma and Singh 2019; Alamdari et al., 2020). Due to the existing challenges faced by the traditional RSs such as cold-start, scalability, and accuracy, researchers have also been exploring and studying new approaches such as deep learning (machine learning technique) for RSs in recent years (Batmaz et al., 2019).

According to a study by Alamdari et al. (2020), since there is an enormous amount of information produced by e-commerce platforms in this era, RS also plays a crucial role in tackling the information overload issue. Some studies suggest that product RS could heterogenized B2C (Business to Consumer) sales across a variety of products by recommending different products to consumers, (Hinz, Eckert and Skiera, 2011). Moreover, RSs could improve consumer perceptions of usefulness, promote positive attitudes, and lead to higher use intentions (Li and Karahanna, 2015).

Nowadays, countless applications and platforms, including e-commerce (Shopee, Lazada), social media (Instagram, Twitter), video-sharing (Tiktok, Youtube), health, e-learning, music, Internet of Things (IoT), food industry, marketing, and many more have been widely adopting the RS (Jiawei et al., 2023) to enhance their platforms. Besides, RSs are also frequently implemented in search engines to index and find non-conventional data (Gupta and Pandey, 2019). To sustain a competitive advantage in the digital market, global companies including Amazon, Netflix, Google, and Samsung are constantly updating and offering their recommendation services (Kim, Choi and Li, 2021). For example, Google recommends news according to the users' locations and interests in real-time (Das et al., 2007). On the other hand, Netflix has utilised the state-of-the-art deep neural network approach to build a personalisation recommendation for their users (Bennett and Lanning, 2007), and it proves the effectiveness of RS by becoming the leader among online movies and dramas platforms. Therefore, RS could bring positive impacts on a business' scaling and growth.

2.2. Analysis of selected tool with any other relevant tools

Tools comparison	**Remark**	**Jupyter Notebook**	**Pycharm Community**	**Spyder**
Type of licence and open source licence	State all types of licence	BSD licence	Apache 2 licence	MIT licence
Year founded	When is this tool being introduced?	2014	2010	2009
Founding company	Owner	Jupyter	Jet Brains	Spyder Project Contributor
Licence Pricing	Compare the prices if the licence is used for development and business/commercialization	No additional licence pricing.	No additional licence pricing.	No additional licence pricing.
Supported features	What features does it offer?	Supports over 100 programming languages. Able to display the output of running code cells (plot). Easily re-run and move individual snippets. .ipynb files as notebook code files In-browser editing for code, with indentation, automatic syntax highlighting. combines code, graphics, visualisations, and text in shareable notebooks that run in a web browser.	Supports mainly for pure Python development. Equipped with Intelligent Code Editor, provide language-aware code completion, error detection, and on-the-fly code fixes. Smart search to jump to any class, file or symbol. Code refactorings features with safe rename and delete, extract method, introduce variable, inline variable or method, and other refactorings.	Allows users to run multiple IPython consoles. Allows users to work on multiple development projects at the same time. Includes a powerful debugger. Contains an integrated documentation browser.
Common applications	In what areas is this tool usually used?	Data analysis Machine learning Data Visualization	Web development Scientific programming: Data analysis Machine learning	Interactive testing Data inspection Data analysis Debugging
Customer support	How the customer support is given, e.g. proprietary, online community, etc.	Online community is provided (forum)	Support page, user forum and email support	Online customer service is provided
Limitations	The drawbacks of the software	The cells are run by one No code-style correction No IDE integration	Steep learning curve Limited language support, focus Python Resource-intensive Large installation size	Could not load large datasets Lacks a simple interface for cooperative development

2.3. Justify why the selected tool is suitable

Jupyter Notebook is an excellent choice to develop the product recommender system using python due to its support on in-browser editing for code and output displaying. Its .ipynb files (Jupyter's python notebook code files) allows us to write code snippets that could be run individually, thus we can re-run or run in any order to do easy testing and experiments. It supports text or visual output that will be displayed at every code cell we run, which enhances the code evaluation and debugging process. The notebook code file also supports markdown cells other than code cells, allowing us to write readable code explanations or comments with different headings and hyperlinks, and even to display imported images. These features allow us to do coding and debugging efficiently.

Besides, Jupyter Notebook also provides simple and usable GUI which offers useful menu features, interactive widgets and shortcut keys. For example, we could move the code cell using the up and down button on the menu bar, or insert a new code cell with the shortcut key 'A' or 'B'. Also, we could control the kernel easily, such as interrupting, restarting, clearing the output, reconnecting, or shutting down the kernel. Moreover, the directory navigation is also GUI based in Jupyter Notebook, so we can move around our system's directories at ease. These features enhance the development experience.

Another reason why Jupyter Notebook is a good choice for developing a product recommender system is its ability to combine code, graphics, visualisations, and text in shareable notebooks that run in a web browser. With it, the team could easily share code and results for collaboration. Furthermore, Jupyter Notebook also supports the visualisation of various data formats and sources, including JSON and CSV files. Additionally, its large user base and online resources also make it a suitable choice for the project. By using it, we can prototype and fine-tune our algorithms quickly.

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 products.

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	Sales rank information that indicates how the product is selling compared to other items within the same category.
also_view	string	A product list that indicates the user who viewed the product also viewed the list of products.
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)

The algorithm or approach that I have chosen to imply for the product recommender system is the Simple Recommender System (SRS) . It is a basic type of recommender system that will suggest the top 20 popular products based on the `scores` calculated from the `rating_average` and `rating_count`. Besides, this simple recommender system will also provide users with a filter feature to search the top 20 popular products based on these features entered by the users:

- `product_name`
- `main_category`
- `brand_or_author`
- `price` *(price range)*

Importing Libraries

The first step is to import the libraries. The libraries used for this SRS are:

- pandas
- html
- warnings

File Reading and Features Engineering: products

After importing the electronic_products.json file as a '`products`' dataframe, I noticed that it has 104802 rows (records) and 19 columns (features). Carrying on, the features engineering was done to first drop the unnecessary features, while keeping only the asin as product_id, title as `product_name`, brand as `brand_or_author`, price as `price`, and main_cat as `main_category`. Then, the duplicated records were removed, resulting in a cleansed data of 74434 records with 5 columns. After that, all the data in each feature is analysed, and the dirty data is transformed to meaningful data.

File Reading and Features Engineering: ratings

Next, I import the user_ratings.csv as a '`ratings`' dataframe. It has 7824482 records and 4 columns (features). The features are imported as `user_id`, `product_id`, `rating`, and `timestamp`. Then, in the features engineering process, only `product_id` and `rating` are keeped, while the other two features are dropped. From the `rating` features, I create two new features, which are the `rating_average` and the `rating_count` for each unique product.

Merge products and ratings into products_merge

The products and ratings, where they already undergo the features engineering processes, are merged into a new dataframe, '`products_merge`'. The products' records are all remained,

therefore, the 'products_merge' dataframe has 74434 records and 7 columns (features). The null value of the rating_average and the rating_count are all replaced by 0.

Simple Recommender System (SRS)

A new 'score' feature on a scale 1-5, which represents the popularity of a product, is calculated using the weighted rating formula provided by IMDB, based on the rating_average and the rating_count. The formula of the weighted rating is as follows:

$$Weighted Rating(\mathbf{WR}) = \left(\frac{\mathbf{v}}{\mathbf{v}+\mathbf{m}} \cdot \mathbf{R}\right) + \left(\frac{\mathbf{m}}{\mathbf{v}+\mathbf{m}} \cdot \mathbf{C}\right)$$

- v is the number of rating for a product (represented by rating_count)
- m is the **minimum rating count** required to be listed in the chart (to be calculated)
  - I have considered the 90th percentile of the rating_count.
  - For a product to be recommended, it must have more votes than at least 90% of the products.
- R is the average rating of a product (represented by rating_average)
- C is the **mean of rating average** across the whole dataframe (to be calculated)

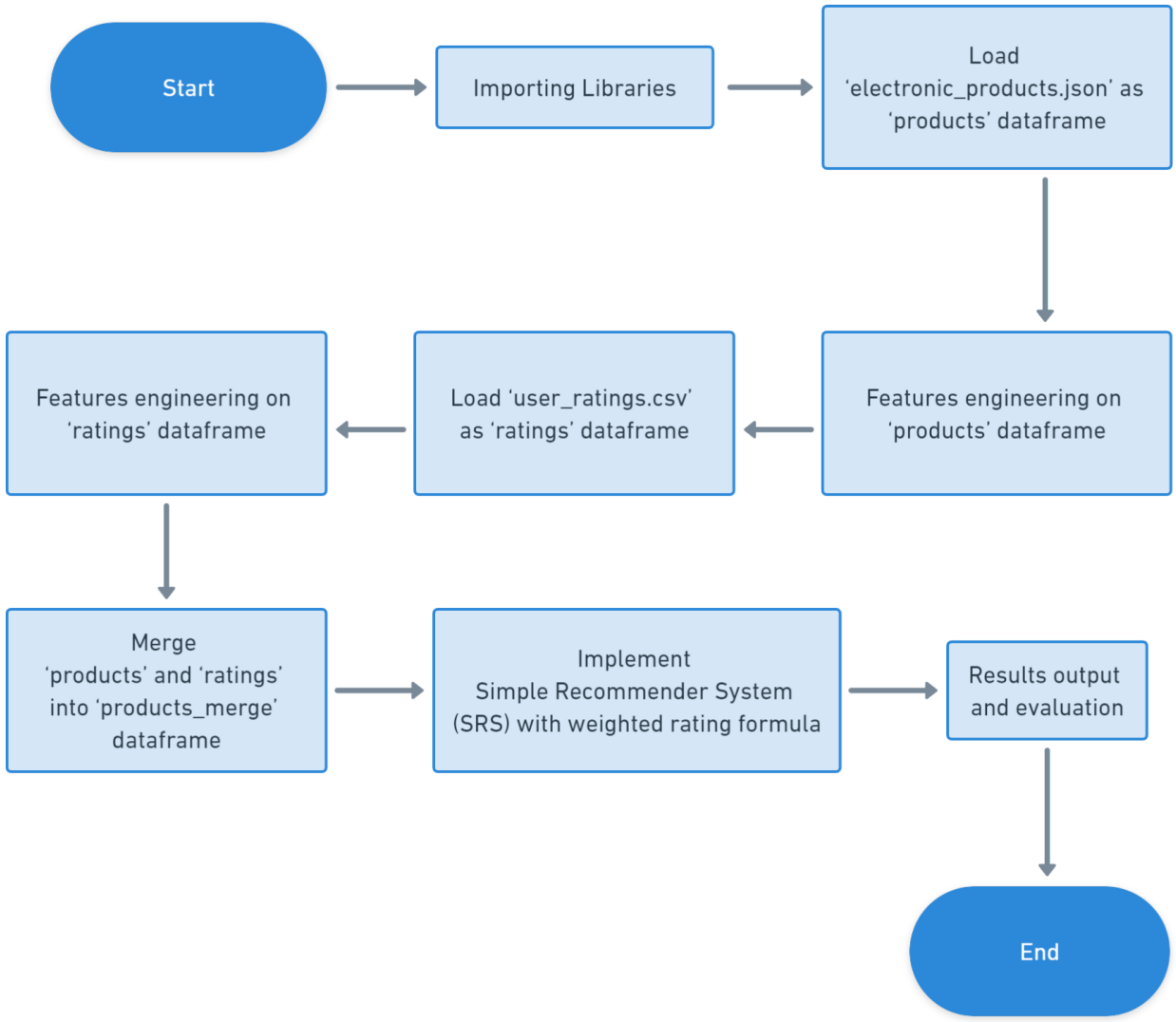
Finally, with the new 'score' feature, the SRS can now output:

- The top 20 most popular products

Furthermore, I have also added filter options on some of the features, and they will output results as below based on users' input on a certain feature:

- The top 20 most popular products based on product_name
- The top 20 most popular products based on main_category
- The top 20 most popular products based on brand_or_author
- The top 20 most popular products based on price range (min price and max price)

3.3. System flowchart/activity diagram



3.4. Proposed test plan/hypothesis

Test Data

After implementing the Simple Recommender System (SRS), the results of the 5 outputs will be tested with appropriate data:

- The top 20 most popular products
  - Input data: (no input)
- The top 20 most popular products based on `product_name`
  - Input data: `speaker`
- The top 20 most popular products based on `main_category`
  - Input data: `camera`
- The top 20 most popular products based on `brand_or_author`
  - Input data: `microsoft`
- The top 20 most popular products based on `price` range (min price and max price)
  - Input data (min price): `10`
  - Input data (max price): `100`

Test Result Evaluation

Also, to evaluate the result of the prototype, since each user's real thoughts and preferences are not easily accessible, and are not solid data, A user feedback survey using **Google Form** will be conducted. Through the survey, the performance and the recommendations results produced by the SRS could be evaluated and compared with the other two modules of the team, and we can know whether the 'trendy' or 'popular' results produced matches the users' preferences or not.

Hypothesis

- **Hypothesis 1**: More than or equal to 50% of users finds 80% of the SRS results match their preferences.
- **Hypothesis 2**: Less than 50% of users find 80% the SRS results match their preferences.

4. Result

4.1. Results

4.1.1 Recommending Top 20 Products

5.3.1 Recommending Top 20 Products

- Sort q_prouncts in descending order based on the score feature column.
- Output the product_name, main_category, brand_or_author, price, rating_average, rating_count, and weighted rating (score) of the top 20 products.

```
# Sort products based on 'score' and recommend the top 20 products
top_20_proucts = q_products.sort_values('score', ascending = False).head(20).reset_index()

columns = ['product_id', 'product_name', 'main_category','brand_or_author', 'price',  'rating_average', 'rating_count', 'score']
top_20_proucts = top_20_proucts [columns]

top_20_proucts.index = top_20_proucts.index + 1

top_20_proucts
```

	product_id	product_name	main_category	brand_or_author	price	rating_average	rating_count	score
1	B000LRMS66	garmin portable friction mount - frustration f...	cell phones & accessories	garmin	18.50	4.756627	8715.0	4.752027
2	B000053HH5	canon ef 70-200mm f/4l usm telephoto zoom lens...	camera & photo	canon	45.12	4.841499	347.0	4.727661
3	B00007GQLU	canon ef 85mm f/1.8 usm medium telephoto lens ...	camera & photo	canon	35.86	4.787934	547.0	4.716221
4	B000I1X3W8	canon ef 70-200mm f/4 l is usm lens for canon ...	camera & photo	canon	59.82	4.869565	253.0	4.714665
5	B000053HC5	canon ef 135mm f/2l usm lens for canon slr cam...	camera & photo	canon	66.78	4.945783	166.0	4.710327
6	B00009UT9B	pelican 1450 case with foam (silver)	camera & photo	pelican	124.95	4.786307	482.0	4.705459
7	B00006I53X	canon ef 70-200mm f/2.8l is usm telephoto zoom...	camera & photo	canon	958.00	4.839080	261.0	4.691095
8	B000CKVOOY	arkon folding tablet stand for ipad air ipad m...	computers	arkon	12.95	4.707955	1873.0	4.687381
9	B000092TT0	polk audio psw505 12-inch powered subwoofer (s...	home audio & theater	polk audio	7.69	4.720169	947.0	4.679673
10	B00020M1U0	sanus vmpl50b vision mount tilting mount for 3...	home audio & theater	twowings	0.00	4.767442	387.0	4.669008
11	B00004XOM3	canon ef 100mm f/2.8 macro usm fixed lens for ...	camera & photo	canon	49.23	4.794613	297.0	4.666549
12	B0002JY712	panasonic cordless telephone battery (hhr-p104a)	office products	panasonic	13.09	4.711515	825.0	4.665444
13	B000NP3DJW	canon speedlite 580ex ii flash for canon eos d...	camera & photo	canon	189.98	4.749441	447.0	4.664463
14	B000PMGZDO	crucial 4gb kit (2gbx2) ddr2 667mhz (pc2-5300)...	all electronics	crucial	1.54	4.734375	448.0	4.650344
15	B00006I53W	canon ef 70-200mm f/2.8l usm telephoto zoom le...	camera & photo	canon	58.60	4.859551	178.0	4.648328
16	B00005LEN4	nikon af fx nikkor 50mm f/1.8d lens for nikon ...	camera & photo	nikon	11.99	4.665763	1107.0	4.632150
17	B0007WK8KS	canon ef-s 60mm f/2.8 macro usm fixed lens for...	camera & photo	canon	25.97	4.815920	201.0	4.630943
18	B0002Y5WZM	canon 2400 slr gadget bag for eos slr cameras	camera & photo	canon	0.00	4.662780	1115.0	4.629466
19	B000JE7GPY	belkin 6-outlet pivot-plug wall mount power st...	home audio & theater	belkin	2.00	4.628305	1929.0	4.609300
20	B0002Y5WZC	canon 200dg digital camera gadget bag -black	camera & photo	canon	49.95	4.676349	482.0	4.600716

4.1.2 Recommending Top 20 Products According to product_name

5.3.2 Recommending Top 20 Products According to product_name

```
# E.g. speaker
product_name = input("Enter the product name : ")
```

```
Enter the product name : speaker
```

```
top_20_product_name = q_products[q_products['product_name'].str.contains(product_name.lower())]

top_20_product_name = top_20_product_name.sort_values('score', ascending = False).reset_index()[columns]
top_20_product_name.index = top_20_product_name.index + 1

top_20_product_name.head(20)
```

	product_id	product_name	main_category	brand_or_author	price	rating_average	rating_count	score
1	B00005T3C8	polk audio rc65i 2-way premium in-wall 6.5" sp...	home audio & theater	polk audio	7.67	4.694656	262.0	4.559308
2	B0002WPSBC	logitech z-5500 thx-certified 5.1 digital surr...	computers	logitech	0.00	4.590796	804.0	4.547056
3	B00005T3BD	polk audio rc60i 2-way premium in-ceiling 6.5"...	home audio & theater	polk audio	3.78	4.627848	395.0	4.539289
4	B0002SQ2P2	logitech z-2300 thx-certified 2.1 speaker syst...	all electronics	logitech	0.00	4.551916	1435.0	4.527733
5	B000OG6I6A	sony ss-b3000 bookshelf speakers (pair, black)	home audio & theater	sony	0.00	4.616188	383.0	4.525706
6	B0001VGFKW	yamaha ns-aw150bl 2-way indoor/outdoor speaker...	home audio & theater	yamaha audio	4.37	4.576923	546.0	4.513969
7	B00006JPDI	bic america dv62si bookshelf speakers (pair, b...	home audio & theater	bic america	8.18	4.634146	246.0	4.496156
8	B0002SQ0A4	logitech x-230 2.1 2-piece dual drive speakers...	all electronics	logitech	0.00	4.552743	474.0	4.481852
9	B000JNA4LS	harman kardon go + play portable speakers syst...	home audio & theater	harman kardon	0.00	4.693333	150.0	4.471047
10	B000OG4E1G	loopilops bluetooth soundbar audio tv speaker...	home audio & theater	cowin	2.00	4.558642	324.0	4.456788
11	B000OG88KY	sony ss-b1000 5-1/4-inch bookshelf speakers (p...	home audio & theater	sony	0.00	4.500752	665.0	4.451324
12	B0002WTK4S	polk audio rm6750 5.1 channel home theater spe...	home audio & theater	polk audio	11.75	4.626437	174.0	4.439203
13	B00009WBYL	bic america dv-62clrs 6-inch 2-way center chan...	home audio & theater	bic america	4.78	4.784946	93.0	4.435574
14	B00011KLOI	bic america venturi dv64 2-way tower speaker, ...	home audio & theater	bic america	7.67	4.738318	107.0	4.434826
15	B0002BEQJ8	metra 70-1761 radio wiring harness for toyota ...	car electronics	metra	4.79	4.515000	400.0	4.433873
16	B0001DBEM4	harman kardon soundsticks ii plug and play mul...	all electronics	harman kardon	0.00	4.476190	651.0	4.426611
17	B000P0PF9G	polk audio db651 6.5"/6.75" 2-way marine certi...	car electronics	polk audio	7.43	4.466667	405.0	4.389189
18	B000FYZARY	jbl l830 3-way high performance 6-inch bookshe...	home audio & theater	jbl	0.00	4.810811	74.0	4.387369
19	B00067OLOS	fluance sxbp2 home theater bipolar surround so...	home audio & theater	fluance	9.47	4.725275	91.0	4.382280
20	B000P0PF8C	polk audio db521 5.25-inch coaxial speakers (p...	car electronics	polk audio	7.43	4.620370	108.0	4.340623

4.1.3 Recommending Top 20 Products According to main_category

5.3.3 Recommending Top 20 Products According to main_category

```
# e.g camera
main_category = input("Enter the main category : ")
```

```
Enter the main category : camera
```

```
top_20_main_category = q_products[q_products['main_category'].str.contains(main_category.lower())]

top_20_main_category = top_20_main_category [columns].sort_values('score', ascending = False).reset_index()
top_20_main_category.index = top_20_main_category.index + 1

top_20_main_category.head(20)
```

	index	product_id	product_name	main_category	brand_or_author	price	rating_average	rating_count	score
1	7115	B000053HH5	canon ef 70-200mm f/4l usm telephoto zoom lens...	camera & photo	canon	45.12	4.841499	347.0	4.727661
2	15388	B00007GQLU	canon ef 85mm f/1.8 usm medium telephoto lens ...	camera & photo	canon	35.86	4.787934	547.0	4.716221
3	60041	B000I1X3W8	canon ef 70-200mm f/4 l is usm lens for canon ...	camera & photo	canon	59.82	4.869565	253.0	4.714665
4	7117	B000053HC5	canon ef 135mm f/2l usm lens for canon slr cam...	camera & photo	canon	66.78	4.945783	166.0	4.710327
5	19194	B00009UT9B	pelican 1450 case with foam (silver)	camera & photo	pelican	124.95	4.786307	482.0	4.705459
6	13386	B00006I53X	canon ef 70-200mm f/2.8l is usm telephoto zoom...	camera & photo	canon	958.00	4.839080	261.0	4.691095
7	5267	B00004XOM3	canon ef 100mm f/2.8 macro usm fixed lens for ...	camera & photo	canon	49.23	4.794613	297.0	4.666549
8	70340	B000NP3DJW	canon speedlite 580ex ii flash for canon eos d...	camera & photo	canon	189.98	4.749441	447.0	4.664463
9	13422	B00006I53W	canon ef 70-200mm f/2.8l usm telephoto zoom le...	camera & photo	canon	58.60	4.859551	178.0	4.648328
10	8155	B00005LEN4	nikon af fx nikkor 50mm f/1.8d lens for nikon ...	camera & photo	nikon	11.99	4.665763	1107.0	4.632150
11	37996	B0007WK8KS	canon ef-s 60mm f/2.8 macro usm fixed lens for...	camera & photo	canon	25.97	4.815920	201.0	4.630943
12	31885	B0002Y5WZM	canon 2400 slr gadget bag for eos slr cameras	camera & photo	canon	0.00	4.662780	1115.0	4.629466
13	31890	B0002Y5WZC	canon 200dg digital camera gadget bag -black	camera & photo	canon	49.95	4.676349	482.0	4.600716
14	46049	B000BNY64C	stk canon bp-511 bp-511a battery - 2 pack 2200...	camera & photo	sterlingtek	24.99	4.671260	508.0	4.599553
15	64530	B000KJQ1DG	nikon d40 6.1mp digital slr camera kit with 18...	camera & photo	nikon	0.00	4.660036	553.0	4.594389
16	37987	B0007WK8LC	canon bg-e3 battery grip for eos rebel xti & x...	camera & photo	canon	75.95	4.802326	172.0	4.591644
17	13495	B00006I5J7	nikon lc-52 snap on front lens cap	camera & photo	nikon	0.00	4.677003	387.0	4.583850
18	29102	B0002EXF38	slik complete tripod pro 700dx amt tripod with...	camera & photo	slik	12.61	4.734513	226.0	4.575849
19	51319	B000EOSHGQ	nikon af-s vr micro-nikkor 105mm f/2.8g if-ed ...	camera & photo	nikon	63.82	4.739336	211.0	4.570052
20	43692	B000AO3L84	canon 430ex speedlite flash for canon eos slr ...	camera & photo	canon	32.87	4.707510	253.0	4.566650

4.1.4 Recommending Top 20 Products According to brand_or_author

5.3.4 Recommending Top 20 Products According to brand_or_author

```
# e.g. microsoft
brand_or_author = input("Enter the brand or author : ")
```

```
Enter the brand or author : microsoft
```

```
top_20_brand_or_author = q_products[q_products['brand_or_author'].str.contains(brand_or_author.lower())]

top_20_brand_or_author = top_20_brand_or_author [columns].sort_values('score', ascending = False).reset_index()
top_20_brand_or_author.index = top_20_brand_or_author.index + 1

top_20_brand_or_author.head(20)
```

	index	product_id	product_name	main_category	brand_or_author	price	rating_average	rating_count	score
1	9390	B00005TQ08	microsoft intellimouse optical mouse	all electronics	microsoft	1.98	4.514894	235.0	4.382093
2	11884	B00006B7HB	microsoft wheel mouse optical	all electronics	microsoft	0.00	4.403315	181.0	4.248596
3	7244	B00005853Z	microsoft trackball explorer	all electronics	microsoft	8.82	4.417808	146.0	4.229189
4	4100	B00004S9AK	microsoft d58-00002 intellimouse optical	all electronics	microsoft	0.00	4.436508	126.0	4.219748
5	28642	B0002CPBUK	microsoft digital media pro keyboard	all electronics	microsoft	0.00	4.374269	171.0	4.215191
6	72455	B000OY71LS	microsoft 15.6" neoprene laptop sleeve	all electronics	microsoft	0.00	4.511364	88.0	4.205020
7	10213	B0000642RX	microsoft natural keyboard elite	all electronics	microsoft	4.27	4.240084	479.0	4.184815
8	21235	B0000AOWVN	microsoft natural multimedia keyboard	all electronics	microsoft	2.72	4.339450	109.0	4.112498
9	27648	B00025O7FC	microsoft wheel optical mouse	all electronics	microsoft	9.75	4.246377	138.0	4.073841
10	67800	B000MKKTKE	microsoft impact messenger bag for 17.3" lapto...	computers	samsill/microsoft	0.00	4.347222	72.0	4.030856
11	3670	B00002JXFH	microsoft natural keyboard pro	all electronics	microsoft	0.00	4.338710	62.0	3.987932
12	56733	B000GE9XQ2	microsoft lifecam vx-1000	all electronics	microsoft	12.87	4.103261	184.0	3.985395
13	72456	B000OY71MM	microsoft 17" neoprene laptop sleeve	all electronics	microsoft	0.00	4.526316	38.0	3.967132
14	57428	B000GOUE7O	microsoft intellimouse explorer 3.0 optical mo...	all electronics	microsoft	0.00	4.077922	154.0	3.943608
15	42919	B000A6PPOK	microsoft natural ergonomic keyboard 4000	all electronics	microsoft	0.00	3.950495	2828.0	3.943184
16	9389	B00005TQ0A	microsoft wheel mouse optical	all electronics	microsoft	0.00	4.620690	29.0	3.923815
17	13773	B00006IJO4	microsoft broadband networking wireless base ...	all electronics	microsoft	2.10	4.263158	57.0	3.913113
18	11274	B000068MP2	microsoft optical mouse (blue)	all electronics	microsoft	0.00	4.136364	88.0	3.910377
19	42908	B000A6NUU6	microsoft notebook optical mouse 3000	all electronics	microsoft	3.43	4.000000	189.0	3.896536
20	42897	B000A6LSKU	microsoft m03-00090 wireless optical mouse 500...	all electronics	microsoft	0.00	3.994286	175.0	3.884232

4.1.5 Recommending Top 20 Products According to price range

5.3.5 Recommending Top 20 Products According to price range

```
# e.g. 10, 100
min_price = int(input("Enter the minimum price : "))
max_price = int(input("Enter the maximum price : "))
```

```
Enter the minimum price : 10
Enter the maximum price : 100
```

```
top_20_within_price_range = q_products[(q_products['price'] >= min_price) & (q_products['price'] <= max_price)]

top_20_within_price_range = top_20_within_price_range [columns].sort_values('score', ascending = False).reset_index()
top_20_within_price_range.index = top_20_within_price_range.index + 1

top_20_within_price_range.head(20)
```

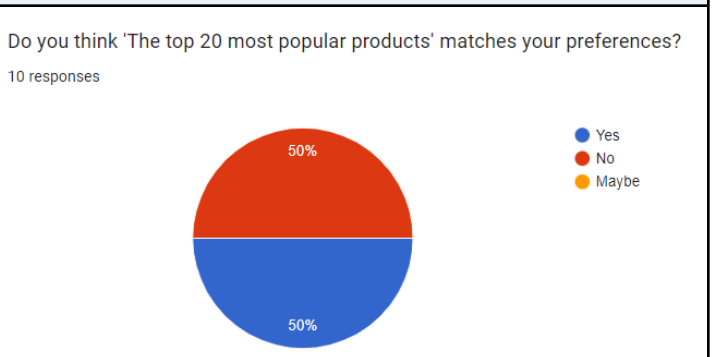
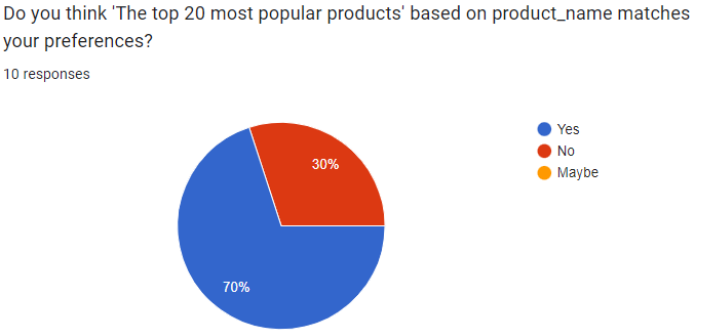
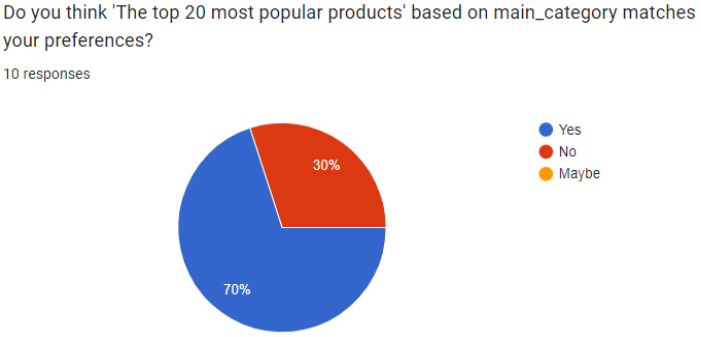
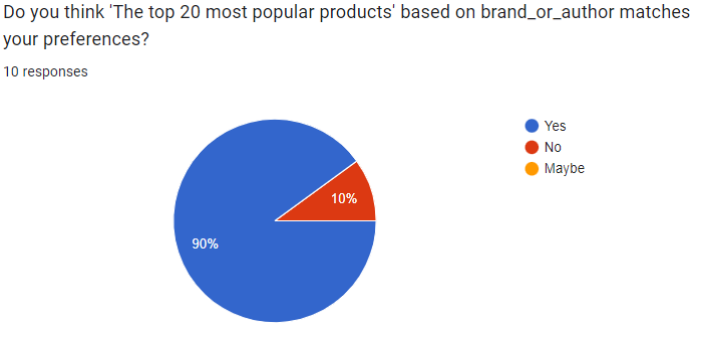
	index	product_id	product_name	main_category	brand_or_author	price	rating_average	rating_count	score
1	66353	B000LRMS66	garmin portable friction mount - frustration f...	cell phones & accessories	garmin	18.50	4.756627	8715.0	4.752027
2	7115	B000053HH5	canon ef 70-200mm f/4l usm telephoto zoom lens...	camera & photo	canon	45.12	4.841499	347.0	4.727661
3	15388	B00007GQLU	canon ef 85mm f/1.8 usm medium telephoto lens ...	camera & photo	canon	35.86	4.787934	547.0	4.716221
4	60041	B000I1X3W8	canon ef 70-200mm f/4 l is usm lens for canon ...	camera & photo	canon	59.82	4.869565	253.0	4.714665
5	7117	B000053HC5	canon ef 135mm f/2l usm lens for canon slr cam...	camera & photo	canon	66.78	4.945783	166.0	4.710327
6	48068	B000CKVOOY	arkon folding tablet stand for ipad air ipad m...	computers	arkon	12.95	4.707955	1873.0	4.687381
7	5267	B00004XOM3	canon ef 100mm f/2.8 macro usm fixed lens for ...	camera & photo	canon	49.23	4.794613	297.0	4.666549
8	30238	B0002JY712	panasonic cordless telephone battery (hhr-p104a)	office products	panasonic	13.09	4.711515	825.0	4.665444
9	13422	B00006I53W	canon ef 70-200mm f/2.8l usm telephoto zoom le...	camera & photo	canon	58.60	4.859551	178.0	4.648328
10	8155	B00005LEN4	nikon af fx nikkor 50mm f/1.8d lens for nikon ...	camera & photo	nikon	11.99	4.665763	1107.0	4.632150
11	37996	B0007WK8KS	canon ef-s 60mm f/2.8 macro usm fixed lens for...	camera & photo	canon	25.97	4.815920	201.0	4.630943
12	31890	B0002Y5WZC	canon 200dg digital camera gadget bag -black	camera & photo	canon	49.95	4.676349	482.0	4.600716
13	46049	B000BNY64C	stk canon bp-511 bp-511a battery - 2 pack 2200...	camera & photo	sterlingtek	24.99	4.671260	508.0	4.599553
14	37987	B0007WK8LC	canon bg-e3 battery grip for eos rebel xti & x...	camera & photo	canon	75.95	4.802326	172.0	4.591644
15	50327	B000EEZCEG	sandisk 1 gb microsd card (sdsdq-1024-a10m, us...	all electronics	sandisk	19.99	4.721774	248.0	4.577067
16	29102	B0002EXF38	slik complete tripod pro 700dx amt tripod with...	camera & photo	slik	12.61	4.734513	226.0	4.575849
17	51319	B000EOSHGQ	nikon af-s vr micro-nikkor 105mm f/2.8g if-ed ...	camera & photo	nikon	63.82	4.739336	211.0	4.570052
18	43692	B000AO3L84	canon 430ex speedlite flash for canon eos slr ...	camera & photo	canon	32.87	4.707510	253.0	4.566650
19	47173	B000BY52NK	nikon d200 10.2mp digital slr camera (body onl...	camera & photo	nikon	66.78	4.754011	187.0	4.563802
20	38700	B00080G0BK	thermaltake mobile fan ii adjustable speed ext...	all electronics	thermaltake	13.99	4.617785	641.0	4.562349

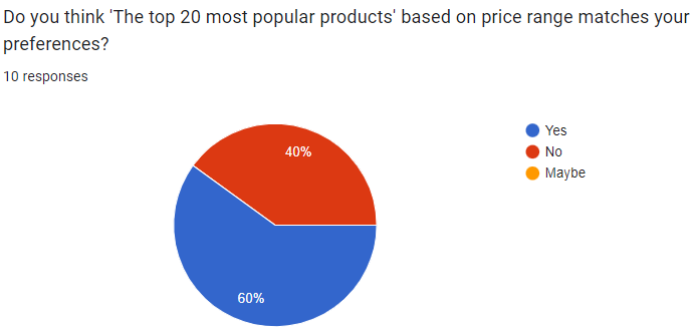
4.2. Discussion/Interpretation

Results Evaluation

Results	Discussions
The top 20 most popular products	Successfully recommended 20 products descendingly with the highest score first.
The top 20 most popular products based on `product_name` • Input data: `speaker`	Successfully recommended 20 products descendingly with the highest score first according to the 'speaker' filter in product_name.
The top 20 most popular products based on `main_category` • Input data: `camera`	Successfully recommended 20 products descendingly with the highest score first according to the 'camera' filter in main_category.
The top 20 most popular products based on `brand_or_author` • Input data: `microsoft`	Successfully recommended 20 products descendingly with the highest score first according to the 'microsoft' filter in brand_or_author.
The top 20 most popular products based on `price` range (min price and max price) • Input data (min price): `10` • Input data (max price): `100`	Successfully recommended 20 products descendingly with the highest score first according to the price range filter from minimum price of 10 to maximum price of 100.
Interpretation: The Simple Recommender System (SRS) has effectively provided recommendations according to other users' ratings as a popularity metrics on the products. It is a good approach to recommend trendy products to newly registered users or users who don't have purchase history. WIth the recommendations, they could get some ideas on which electronic products are popular, useful, and should be purchased.	

Google Form Results (User Feedback)

Results	User Feedback
The top 20 most popular products	
The top 20 most popular products based on `product_name` • Input data: `speaker`	
The top 20 most popular products based on `main_category` • Input data: `camera`	
The top 20 most popular products based on `brand_or_author` • Input data: `microsoft`	

The top 20 most popular products based on `price` range (min price and max price) • Input data (min price): `10` • Input data (max price): `100`	
Interpretation: Based on all the pie charts' summary results above, the user feedback shows more than or equal to 50% of users agreeing that 100% of the SRS results match their preferences. Therefore, Hypothesis 1 is accepted. Thus, we can conclude that the Simple Recommender System has recommended popular electronic products that match the users' preferences successfully.	

5. Discussion and Conclusion

5.1. Achievements

With this Simple Recommender System (SRS) module, it has successfully produced a functioning recommendation system which will recommend 20 trendy products based on other users' ratings to users, especially to the new users who have not yet made any purchases on the site. Also, this module has implemented four extra filter features, allowing users to search for the relevant popular products by entering either the product's name, main category, brand or author, or price range. The provided filter options are useful if the users wish to narrow down the product discovery scope. With it, users could find or discover their preferred products efficiently.

In conclusion, the project's objectives are said to be accomplished with this SRS module.

Firstly, it has proved to improve the efficiency metric of **system quality** where users are presented with a user-friendly feature that enhances the product discovery process. SRS can help users in discovering trendy products quickly, which has brought additional value to the customers by introducing them with popular product lists, which they might not be able to discover otherwise. Besides, the implementation of the filter features has successfully made the SRS module a more versatile one, allowing users to customise their searches according to their preferences. Therefore, SRS enhances the system quality with its ability to provide popular recommendation and extra control on customers' product discovery experience, which simplifies the process and saves users' time.

Secondly, it has proved to increase the **customer lifetime value (CLV)** by recommending products that match the customer's preferences. Based on the user feedback from the last 'Results' section, users are satisfied with the SRS's recommendations, and this is a significant sign that will increase the customer loyalty, retention, and repeat purchases. The improvement of average number of purchases or purchase frequency will then improve the CLV, and foster the customer base for MyElectronic.

Lastly, the third objective, which is to encourage **upselling and cross-selling**, is also achievable by the SRS module. SRS will recommend popular products, in other words, it is a digital endorsement engine of the trending products of MyElectronic that works based on the users' rating data. If a product is useful and popularly favoured, SRS will automatically boost the product up to the recommendation lists and present it to the users without a salesman or admin working in behind. With the high ratings by other users, the users who have come across with the recommendations will be interested in buying the product. Thus, this is how upselling and cross-selling could be promoted with the SRS.

5.2. Limitations and Future Works

The first limitation is that the current system (SRS) is **only a prototype** and has not yet integrated with the existing e-commerce system of MyElectronic. Due to that, the real output and final results on the production system cannot be visualised. Besides, we cannot guarantee that there are no bugs or errors in the system. To improve the system in the future, we plan to integrate the SRS with MyElectronic's web applications with a visually appealing and easy-to-use interface which will attract more users to the system.

Besides, another limitation of SRS is that it was developed using an **outdated Amazon dataset from 2018**. Also, due to resource limitations (limited RAM space), only a **small portion of the data** could be used to develop the SRS. The amount of records used are less than 100k, while a real e-commerce system could have millions of data records. The lack of data can affect the effectiveness of our recommendation system and lower the accuracy of the results. To solve the issue of using outdated data, we could access the most current and relevant datasets to our project, in this case, are the real production datasets from MyElectronic's application. We could also implement a mechanism for regularly updating the dataset used by the SRS to ensure that it reflects the latest trends and user preferences. On the other hand, we plan to.upgrade the hardware of the system, such as increasing the RAM capacity for more data records to be processed in the development. Additionally, we could also consider using cloud computing services that offer more resources to handle large amounts of data.

Finally, another limitation of SRS is that the **user preferences may change from time to time** according to the evolving latest trend. If SRS is not constantly updated, it could affect the user's satisfactions and feedback as they may rate the same electronic product differently after a period of time. For future improvements, we could collect new or updated feedback from users and incentivise them to provide new feedback by offering rewards. Furthermore, we could manipulate the SRS with a new feature to let it automatically update its recommendation results according to the most recent data in real-time.

Reference & Source

Sources of Datasets

https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews
https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

Development Tools

Jupyter Notebook - web-based interactive computing platform

References

Alamdari, P.M., Navimipour, N.J., Hosseinzadeh, M., Safaei, A.A. and Darwesh, A. (2020) 'A systematic study on the recommender systems in the E-commerce.' *Ieee Access*, 8, pp.115694-115716. doi: 10.1109/ACCESS.2020.3002803.

Batmaz, Z., Yurekli, A., Bilge, A., et al. (2019) 'A review on deep learning for recommender systems: challenges and remedies.' *Artif Intell Rev*, 52, 1–37. doi: 10.1007/s10462-018-9654-y.

Bennett, J. and Lanning, S. (2007) 'The netflix prize.', *Proceedings of KDD cup and workshop*, Vol. 2007, p. 35.

Das, A.S., Datar, M., Garg, A. and Rajaram, S. (2007) 'Google news personalization: scalable online collaborative filtering', *Proceedings of the 16th international conference on World Wide Web*, pp. 271-280.

Gupta, V. and Pandey, S. (2019) 'Recommender Systems for Digital Libraries: A review of concepts and concerns.' *Library Philosophy and Practice*, 2019.

Hinz, O., Eckert, J. and Skiera, B. (2011). 'Drivers of the long tail phenomenon: An empirical analysis.', *Journal of Management Information Systems*, 27(4), 43-69.

Jannach, D., Pu, P., Ricci, F. and Zanker, M. (2021) 'Recommender Systems: Past, Present, Future.' *AI Magazine*, 42(3), 3-6. doi: 10.1609/aimag.v42i3.18139

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)', *cseweb.ucsd.edu*. [online] Available at: https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/ (Accessed: 1 April 2023).

Jiawei, C., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023) 'Bias and Debias in Recommender System: A Survey and Future Directions.', *ACM Transactions on Information Systems*, 41(3), 67. doi: 10.1145/3564284.

Julian, M. (n.d.) ‘Recommender Systems and Personalization Datasets’, *cseweb.ucsd.edu*. [online] Available at: https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews (Accessed: 1 April 2023).

Jussi K. (2019) ‘A digital bookshelf: original work on recommender systems.’ [online] Available at: https://jussikarlgren.wordpress.com/2017/10/01/a-digital-bookshelf-original-work-on-recommender-systems (Accessed: 29 March 2023).

Kim, J., Choi, I. and Li, Q. (2021) ‘Customer satisfaction of recommender system: Examining accuracy and diversity in several types of recommendation approaches.’, *Sustainability*, 13(11), p.6165. doi: 10.3390/su13116165.

Li, S.S. and Karahanna, E. (2015) ‘Online Recommendation Systems in a B2C E-Commerce Context: A Review and Future Directions,’, *Journal of the Association for Information Systems*, 16(2). doi: 10.17705/1jais.00389.

Milano, S., Taddeo, M. and Floridi, L. (2020) 'Recommender systems and their ethical challenges', *AI & Soc*, 35, 957-967. doi: 10.1007/s00146-020-00950-y.

Raven360 (2023) ‘How to Increase Customer Lifetime Value’, 3 March. [online] Available at: https://www.raven360.com/blog/how-to-increase-customer-lifetime-value (Accessed: 28 March 2023).

Roy, S., Sharma, M. and Singh, S.K. (2019) ‘Movie recommendation system using semi-supervised learning.’, *Global conference for advancement in technology (GCAT)*, 2019, pp. 1-5. IEEE. doi: 10.1109/GCAT47503.2019.8978353.

Saurav, A. (2022) ‘Amazon Product Reviews’, *kaggle*. [online] Available at: https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews (Accessed: 1 April 2023).

Sharma, J., Sharma, K., Garg, K., & Sharma, A. (2021) 'Product Recommendation System: A Comprehensive Review', *IOP Conference Series: Materials Science and Engineering*, 1022, 012021. doi: 10.1088/1757-899X/1022/1/012021.

Sharma, R. and Singh, R. (2016) ‘Evolution of recommender systems from ancient times to modern era: a survey.’ *Indian Journal of Science and Technology*, 9(20), 1-12. doi: 10.17485/ijst/2016/v9i20/88005

Sophia, B. (2022) ‘Cross-Selling and Upselling: The Ultimate Guide’, *HubSpot*, 14 September. [online] Available at: https://blog.hubspot.com/sales/cross-selling (Accessed: 28 March 2023).