EVALUATION OF ACCURACY BETWEEN ITEM-BASED AND MATRIX FACTORIZATION RECOMMENDER SYSTEM

Oloyede I., Hassan I. A.

Department of Statistics, University of Ilorin, Ilorin, Nigeria

Corresponding Author: Oloyede I., oloyede.i@unilorin.edu.ng

ABSTRACT: Recommender systems are used by e-commerce websites and streaming services to predict user opinion about products. This work examined two specific recommender algorithms, matrix factorization collaborative filtering algorithm and Item-based collaborative filtering, which utilizes item similarity. This study is to compare the prediction accuracy of the algorithms using Mean Square Error and Root Mean Square Error criteria. The work yielded a result which indicated that the matrix factorization collaborative filtering algorithm is more accurate than the Item-based collaborative filtering algorithm. From the study, the results of evaluation metrics showed that matrix factorization method having RMSE of 0.916250 and MAE 0.708731 scales slightly better than the item-based method having RMSE and MAE of 0.937089 and 0.719434. The study concluded that the matrix factorization method is more accurate than the item-based method when evaluating their prediction accuracy using RMSE and MAE.
KEYWORDS: item-based, recommendation system, Matrix factorization, filtering.

1. INTRODUCTION

There are different kinds of products or materials that are purchased or used in day-to-day lives. The products have product specifications, the consumer compared the products with similar attribute on the Internet, read the feedback from anonymous users, and then making decision about the product to buy. A recommender system or recommendation engine makes these feasible ([GU15]). Recommendation system is a machine-learning technique to predict what new items a user would like based on associations with the user's previous items and it is widely used in the field of e-commerce applications. And the development of recommender systems is a multi-disciplinary fields which involves experts from various fields such as Statistics, Artificial intelligence, Human Computer Interaction, Information Technology, Data Mining, Mathematics, Decision Support Systems and Marketing etc. Machine Learning is the science of getting computers to learn and act like humans do, by using statistical techniques to give computer systems the ability to learn with data and improve their learning from experience without being explicitly programmed.

It is also a tools and techniques providing suggestions for items to be of use to a user. The suggestions provided are aimed at supporting the users in various decision-making processes, such as what items to buy, what music to listen, or what news to read. Recommender systems have proven to be valuable means for online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce ([FLB11]). It also help in suggestions, such as useful products on e-commerce websites, videos on YouTube, friends' recommendations on Facebook, book recommendations on Amazon, news recommendations on online news websites, and the list goes on. Through this flexible data and behaviour-driven algorithms, businesses can increase conversions by helping to ensure that relevant choices are automatically suggested to the right customers at the right time with cross-selling or up-selling ([Vig13]).

There are different types of recommender systems that vary in terms of the addressed domain, which are collaborative filtering recommender systems, content-based recommender systems, hybrid recommender system, knowledge-based recommender systems etc. ([FLB11]).

Content-based recommender systems is a type of recommendation system, whereby the system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items ([Vig13]). Knowledge-based recommendation system is a type of recommendation systems whereby system recommend items based on specific domain knowledge about how certain item features meet users needs and preferences and ultimately, how the item is useful for the user. Hybrid recommender systems is a type of recommender system that are based on the combination of different recommender techniques. By combining various recommender systems, it can eliminate the disadvantages of one system with the advantages of another system and thus build a more robust system ([FLB11], [GU15]).

Collaborative filtering recommendation system is a type of recommendation system whereby the system recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users. Collaborative filtering is considered to be the most popular and widely implemented technique in recommender system and it is of two type. They are: user-based collaborative filtering recommendation system and item-based collaborative filtering recommendation system ([FLB11]).

User-based collaborative filtering recommendation system is a type of collaborative filtering recommendation system whereby users similar to current user are determined. Based on this user similarity, their interested/used items can be recommended to other users. Item-based collaborative recommendation system is a type of collaborative filtering recommendation system whereby items similar to the items that are being currently used by a user are determined. Based on the item-similarity score, the similar item will be presented to the users for cross-selling and up-selling type of recommendations ([Vig13]).

Tapestry is one of the earliest implementations of collaborative filtering-based recommender systems. It was designed to filter e-mails received from mailing lists and newsgroup postings. In this system, each user could write a comment (annotation) about each e-mail message and share these annotations with a group of users. A user could then filter these e-mail messages by writing queries on these annotations. Tapestry allowed an individual user to benefit from annotations made by other users, the system required an individual user to write complicated queries ([G+92]).The first system to generate automated recommendations was the GroupLens system ([R+94], [K+97]), which provided users with personalized recommendations on Usenet postings. The recommendations for each individual were obtained by identifying a neighbourhood of similar users and recommending the articles that this group of users found useful. Ringo is an online social information filtering system that uses collaborative filtering to build users profile based on their ratings on music albums ([C+08]). Amazon uses topic diversification algorithms to improve its recommendation. The system uses collaborative filtering method to overcome scalability issue by generating a table of similar items offline through the use of item-to-item matrix. The system then recommends other products which are similar online according to the users' purchase history.

Likewise, there are some other technologies that have been applied to recommender systems, including Bayesian networks, clustering, and Horting. Bayesian networks create a model based on a training set with a decision tree at each node and edges representing user information. The model can be built off-line over a matter of hours or days. The resulting model is very small, very fast, and essentially as accurate as nearest neighbour methods ([BHK98]). Clustering techniques work by identifying groups of users who appear to have similar preferences. Once the clusters are created, predictions for an individual can be made by averaging the opinions of the other users in that cluster. Some clustering techniques represent each user with partial participation in several clusters. Horting is a graph-based technique in which nodes are users, and edges between nodes indicate degree of similarity between two users ([A+99]). Predictions are produced by walking the graph to nearby nodes and combining the opinions of the nearby users.

This study employed collaborative filtering recommendation system approach to make a recommender system on the Jumia data. Jumia is an online ecommerce store which sells different types of product ranges from clothes, smartphones, computing devices, household materials etc. And the data was web scraped from the part of their computing product section only and it consist of product name, users rating, user's id and numbers of rating. Web scraping is a process of data extraction from a website. The techniques of item-based collaborative filtering recommendation system and Matrix factorization collaborative filtering recommendation system was compared and the accuracy of this recommender techniques is evaluated by using statistical tools Mean Square Error (MSE) and Root Mean Square Error (RMSE) to check for the accuracy of the recommended items.

3. METHODICAL

Collaborative filtering recommender system uses the preferences of users who have liked similar items in the past. Its success depends on the method of technique used to design it. User similarity is evaluated by matching their preferences on a set of common items. There will be a set of users $U = \{u_1, u_2, \dots, u_p\}$ and a set of items $I = \{i_1, i_2, \dots, i_q\}$ such as books, songs, news articles, gadgets etc. Ratings are stored in a $p \times q$ user-item rating matrix. The ratings from users follow a specified numerical scale indicating the degree of preferences (e.g. 1-bad to 5- excellent).

3.1 Collection of data for creating a user profile

The collection of user data depends on user ratings of an item and that two users are similar on the basis of their ratings on common items e.g., smartphone, but it also depends on other factors such as their background and personal details like age, gender, and user priorities to the use of that product.

3.2 Neighbourhood set generation

In this step, neighbour matrix similarity are set, neighbours are a group of likeminded users of active user. The size of the neighbourhood set could be flexible by choosing the users whose similarity value is above a certain threshold. And a variety of similarity methods can be employed, such as cosine similarity, the Pearson correlation, weight amplification, matrix factorization etc. Pearson correlation coefficient is the most popular method for memory-based collaborative filtering and which is defined by

$$sim(u,v) = \frac{\sum_{i\in C}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i\in C}(r_{u,i} - \bar{r}_i)}^2 \sqrt{\sum_{i\in C}(r_{v,i} - \bar{r}_v)}^2}, \tag{1}$$

where: C is the set of items rated by both users u and v
$r_{u,i}$ is the rating of user u on item m
$\bar{r}_u$ is the average rating of item m

3.3 Prediction and Recommendation

When the neighbours are found, several approaches can be used to combine the ratings of neighbours to compute a prediction value on unseen items for the active user. After predicting how an active user will like specific items which have not been rated yet by the active user, the top-N item set, a set of ordered items with a higher predicted value, is identified and recommended. The predicted rating $pre_{u,i}$, of item i of a user u can be computed by the following formula:

$$pre_{u,i} = r_u + k \sum_{ú\in C} d(u,ú) \times (r_{ú,ı} - \bar{r}_{ú}), \tag{2}$$

where: C denotes the set of neighbours who have rated item
k is a multiplier for normalizing factor
$\bar{r}_{ú}$ is the average rating of user $ú$

3.4 Evaluation of Accuracy

The accuracy of a recommendation algorithm can be evaluated using different types of measurement. Accuracy is the fraction of correct recommendations out of total possible recommendations. The metrics of measuring the accuracy of recommendation filtering systems are divided into statistical and decision support accuracy metrics. This study will look into the statistical accuracy metrics only. And for this study, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the statistical accuracy metrics that are going to be used to do the evaluation of accuracy of the recommender system.

3.5 Mean Absolute Error (MAE)

In statistics, mean absolute error (MAE) is a measure of difference between two continuous variables. Assume X and Y are variables of paired observations that express the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. Consider a scatter plot of n points, where point i has coordinates (xi, yi). Mean Absolute Error (MAE) is the average vertical distance between each point and the Y=X line, which is also known as the One-to-One line. MAE is also the average horizontal distance between each point and the Y=X line. It is used to compute the average of all the absolute value differences between the true and the predicted rating. The lower the MAE, the better is the accuracy. It can range from 0 to infinity, where infinity is the maximum error depending on the rating scale of the measured value. The Mean Absolute Error is determined with the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|d_i| - |\hat{d}_i| \quad (3)$$

where: d_i is the actual rating
$\hat{d}_i$ is the predicted rating
n is the amount of ratings

3.6 Root Mean Square Error (RMSE) / Root Mean Square Deviation (RMSD)

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. The RMSD represents the sample standard deviation of the differences between predicted values and observed values. These individual differences are called residuals when the calculations are performed over the data sample that was used for estimation, and are called prediction errors when computed out-of-sample. The RMSD serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSD is a measure of accuracy, to compare forecasting errors of different models for a particular data and not between datasets, as it is scale-dependent. The root mean square error between the true ratings and predicted ratings is determined with the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(d_i - \hat{d}_i\right)^2} \quad (4)$$

where: d_i is the actual rating
$\hat{d}_i$ is the predicted rating and n is the amount of ratings

4. DATA ANALYSES AND INTERPRETATION

4.1 Intuition behind Using Matrix Factorization

In a recommendation system there is a group of users and a set of items. Given that each user have rated some items in the system, the interest is to predict how the users would rate the items that they have not yet rated, such that the recommendations can be made to the users. In this case, all the information available about the existing ratings can be represented in a matrix. Assume that there are 5 users and 10 items, and ratings are integers ranging from 1 to 5, this can be shown in matrix form in the table 1.

Table 1. The existing ratings ([Alb17])

	D1	D2	D3	D4
U1	5	3	-	1
U2	4	-	-	1
U3	1	1	-	5
U4	1	-	-	4
U5	-	1	5	4

4.2 Recommender System Implementation Using Matrix Factorization

Table 2. Data Frame for the Datasets

	User Id	Item Id	Rating	Product
0	0	50	5	HP 15 Intel Celeron 1.6GHz
1	0	172	5	Waafox Ideabook Air WN1401 Intel X5 8350
2	0	133	1	Excel Plus 32GB MicroSDHC Memory Card
3	196	242	3	Lenovo Ideapad Mini Laptop 110S-11IBR
4	186	302	3	Dell Inspiron 15 7th Gen Touchscreen

Table 2 showed the data frame of the data where user identity (user-id) is the id of a specific user that rate an item (products), while Item identity is the id of an item rate by the user, the rating is the amount of rating giving to an item by the user and the product is the product name of each product corresponding to each item-id, e.g.

HP 15 Intel Celeron 1.6GHz corresponding to the item-id 50, item-id 242 correspond to product Lenovo Ideapad Mini Laptop 110S-11IBR etc.
These data frame above can be formatted in such a way that the ratings matrix to be one row per user and one column per product and which is show below:

Table 3. The ratings matrix

user_id /Product	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	...
2	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
3	0	0	5	2	0	0	0	2	0	0	0	2	0	0	0	0	0	4	0	0	...
4	1	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	...
5	3	0	0	0	0	0	4	0	0	0	0	0	0	0	1	0	0	4	0	3	...

The rating matrix above shows the product and the user-id and each row of the matrix corresponds to a given user, and each column corresponds to a given product. All the zero value that is in the cell means that the user has not seen the product before and he has not given rating to the product. Example, user-id 1 may have not seen the product 1 before and he has not given it rating and for the product 3 which have been seen by the user-id 1, he gives it rating 2 due to maybe he did not like the product or some other thing. Likewise, for the product 9, it was given 1 rating by user id 4 and the same user-id 4 did not give any rating to the product 14 due to reason that he may have not seen the product.
Also, the data can be de-mean (i.e. normalize by each user mean) and convert it from a data frame to an array, which is shown below:

```
[[4.84009714  -0.04990286  -0.04990286  ...,  -0.04990286  -0.04990286  -0.04990286]
 [-0.12924987 -0.12924987  -0.12924987  ...,  -0.12924987  -0.12924987  -0.12924987]
 [-0.05369671 -0.05369671  -0.05369671  ...,  -0.05369671  -0.05369671  -0.05369671]
 [-0.03050729 -0.03050729  -0.03050729  ...,  -0.03050729  -0.03050729  -0.03050729]]
```

Since the ratings matrix is properly formatted and normalized, the singular value decomposition can be done.

4.2.2 Singular Value Decomposition

By dong Singular Value Decomposition, the latent factors that is needed can be chosen and use it to approximate the original ratings matrix and the rating matrix is shown below after the decomposition

```
(array(
[[-4.97801875e-03,   5.86971868e-03,   1.18186843e-02, ...,   2.95139320e-03,  -1.95703358e-03],
 [-1.10526375e-03,  -4.04545890e-03,  -1.16776791e-02...     9.18855171e-04,   2.17034433e-03],
 [9.44963839e-03,   -1.43519545e-02,   3.93638119e-04, ...,  -2.89764529e-03,   2.86504507e-03],
 [-1.05731072e-02,  -6.80807641e-03,   2.63392883e-03, ...,  -4.84286245e-05,  -1.89077440e-03],
 [6.34420788e-03,   -9.45269844e-03,  -2.69929558e-03, ...,  -1.07208981e-02,  -1.88878158e-02]
 [-1.84854679e-02,   1.28388950e-02,  -8.46988257e-03, ...,  -1.89987575e-03,   1.92850979e-02]]),
…
```

This matrix above can be converted to the diagonal matrix form so as to get predictions and the converted diagonal matrix is shown below:

```
array(
[[  38.18581225,     0.,            0.,           ...,     0.,            0.,            0.        ],
 [    0.,          38.62154312,     0.,           ...,     0.,            0.,            0.        ],
 [    0.,           0.,           39.58855276,    ...,     0.,            0.,            0.        ],
                                                  ...,
 [    0.,           0.,            0.,           ...,   54.46932602,     0.,            0.        ],
 [    0.,           0.,            0.,           ...,     0.,          63.41536276,     0.        ],
 [    0.,           0.,            0.,           ...,     0.,            0.,          84.10679346  ]])
```

4.2.3 Making Predictions from the Decomposed Matrices

The product ratings predictions for every user can be made by following the math and matrix multiply $\boldsymbol{U}$, $\boldsymbol{V^T}$ and $\boldsymbol{\Sigma}$ back to get the rank $\boldsymbol{k}$**=50** approximation of $\boldsymbol{R}$ and also the user mean need to add back to get the actual star ratings prediction.

[[4.28886061	0.14305516	-0.1950795	...,	0.03191195	0.05044975	0.08891033]	
[0.74471587	0.16965927	0.33541808	...,	-0.10110207	-0.0540982	-0.14018846]	
[1.81882382	0.45613623	0.09097801	...,	0.01234452	0.01514752	-0.10995596]	...,
[0.61908871	-0.16176859	0.10673806	...,	-0.01336948	-0.0303543	-0.11493552]	
[1.50360483	-0.03620761	-0.16126817	...,	-0.01090407	-0.03864749	-0.16835943]	
[1.99624816	-0.18598715	-0.1564782	...,	-0.00664061	0.12706713	0.28500112]]	

4.2.4 Making Product Recommendations

With the predictions matrix for every user, recommended product can be made for any user by return the product with the highest predicted rating that the specified user hasn't already rated and also, the list of product the user has already rated will be show below as well.

User 1 liked HP Notebook 15-ra002nia Intel Celeron N3060, Microsoft LifeCam HD-3000 Webcam, HP 255 AMD Quad Core, HP 15 Intel Celeron 1.6GHz.
User 1 did not like Lenovo Ideapad Mini Laptop 110S-11IBR, HP 290 G1 Business Desktop
User 1 recommended product is Acer Aspire 3 A315-51-35DQ Intel Core I3-6006U - with predicted rating: 2.61178326612

User 2 liked Universal 64GB USB 2.0 Silver Metal Swivel Flash Memory, i-Life Zed Air H3 Intel Pentinum N4200, Universal 3G Universal Modem 7.2Mbps
User 2 did not like Persuasion (1995), Babe (1995), Carrington (1995)
User 2 recommended product is BOOST Micro SD Memory Card - 16GB - Black - with predicted rating: 4.56932148577

User 3 liked Apple MacBook Pro 15-inch Laptop With Touch Bar, Quick Heal Anti-Virus Pro, HP 2.4GHZ Wireless Optical Mouse, Team Group C153 Flash Drive32GB – Blue
User 3 did not like ZTE Universal Mobile 4G Wifi / Mifi
User 3 recommended product is Lenovo 15.6 Inch Laptop Backpack BM400 Laptop Rucksack - with predicted rating: 2.64590027641

User 4 liked HP 15 Notebook PC, 4gb Ram 500gb Windows 10, Huawei Universal Huawei Mobile 4G LTE MiFi Wifi – White
User 4 did not like Generic High Speed Transmission Micro SD 64GB TF Memory Card
User 4 recommended product is Apple MacBook Pro MLH32LL/A 15-inch Laptop With Touch Bar – with predicted rating: 2.66096207465

User 5 liked HP 255 AMD Quad-Core 500GB 4GB DVD 15.6SCREEN Windows 10 +16Gb Flash, Transcend 16GB USB Flash Drive
User 5 did not like Acer TravelMate 11.6" Laptop - Intel Celeron Quad Core
User 5 recommended product is Apple IMac MNEA2LL/A 27 Inch, 3.5GHz Intel Core I5, 8GB RAM, 1TB Fusion Drive, Silver - with predicted rating: 2.61198757584

User 6 liked Sandisk 64GB Cruzer Blade USB Flash Drive, Universal Laptop RAM DDR3 8GB, HP Pavilion X360 (Intel Core I5-7200U 2.5GHz, 8GB RAM, 1TB HDD)
User 6 did not like T-bao Tbook X7 14.1" Notebook - Silver, Universal 1GB Metal Key Design USB 2.0 Memory Flash Drive Thumb Pen Silver
User 6 recommended product is Dell Inspiron 13-5378x360 Convertible Intel Core I7 - with predicted rating: 2.6607347669

User 7 liked Universal 1GB Metal Key Design USB 2.0, Universal E-Table Foldable Laptop Stand
User 7 did not like Universal Apple MacBook Air Case 11"-transparent, Vivitek D552 DLP PROJECTOR - WHITE, Generic 32GB High Speed Hone Connection USB2.0 Flash Storage Drive U-Disk BK
User 7 recommended product is Seagate 500GB External Hard Disk Drive HDD-USB 3.0 - with predicted rating: 2.6114108657

4.3 Recommendation Using Item-Based Techniques

Table 4. Data Frame for the Datasets

	User Id	Item Id	Rating	Product
0	1	21	3	HP 255 G6 E2-9000e AMD
1	2	122	5	Dell Inspiron 15 Intel Pentium
2	2	243	2	T-bao Tbook X7 14.1" Notebook
3	169	342	4	Sandisk 8GB MicroSDHC Memory Card
4	286	102	4	i-Life Zed Air Intel Celeron

Table 4 showed the data frame of the data where user identity (user id) is the Id of a specific user that rate an item (products), while Item identity is the Id of an item rate by the user, the rating is the amount of rating giving to an item by the user and the product is the product name of each product corresponding to each item-id, e.g. Dell Inspiron 15 Intel Pentium corresponding to the item-id 122, item-id 102 correspond to product i-Life Zed Air Intel Celeron etc.

Table 5. Similarities matrix between the items (Products)

Title / user_id	HP 15 Intel Pentium Quad Core	Lenovo Ideapad 110-15ISK	Hp 255 G6 AMD Quad-Core	Sandisk 8GB MicroSDHC Memory Card	Innjoo Leapbook A100	Canon Refill Ink Set For Pixma 7240 Cartridges	HSDPA 3.75G Universal Modem	…
0	0	0	1	0	0	0	0	…
2	0	2	0	3	0	0	0	…
3	0	0	3	0	0	2	0	…
4	0	0	0	0	1	0	0	…

Table 5 showed the similarities matrix between the product and each cell consist of the rating the user gave to that product. All the zero value that is in the cell means that the user have not seen the product before and he has not given rating to the product. Example, user id 0 may have not seen the product HP 15 Intel Pentium Quad Core before and he has not given it rating and for the product Hp 255 G6 AMD Quad-Core which have been seen by the user id 0, he gives it rating 1 due to maybe he did not like the product or some other thing. Likewise for the product Canon Refill Ink Set For Pixma 7240 Cartridges, it was given 2 ratings by user id 3 and the same user id 3 did not give any rating to Lenovo Ideapad 110-15ISK due to reason that he may have not seen the product.

Table 6. Most rated product

Title	Rating	Number of Ratings
Sandisk 16GB Cruzer Blade Flash Drive - Black	4.359589	125
HP DeskJet 2130 All-in-One Printer- White	3.803536	101
HP 2.4GHZ Wireless Optical Mouse	4.155512	91
Transcend 1TB USB 3.0 StoreJet	4.007890	71
Sandisk 8GB MicroSDHC Memory Card	3.156701	56
Team Group Flash TE90232GN01 32GB	3.656965	49
HP 255 AMD Quad Core	3.441423	43
HP 255 G6 E2-9000e AMD	3.878319	38
HP DeskJet 2620 All-in-One Printer - V1N01C	3.631090	35
Caden 64GB High Speed Hone Connection USB2.0	3.438228	29
ZTE 4G LTE Mobile Internet WiFi Hotspot	4.252381	27

Table 6 showed the average rating and number of ratings of the products that have the most rating from the users. The first eleven movies that have highest number of ratings was selected. And from the list, Sandisk 16GB Cruzer Blade Flash Drive - Black was the highest rated product and have the highest total number of rating (125) followed by HP DeskJet 2130 All-in-One Printer- White with number of ratings 101 till it reaches ZTE 4G LTE Mobile Internet WiFi Hotspot with number of ratings 17. Thus, this indicate that the product that was rated most by the users is Sandisk 16GB Cruzer Blade Flash Drive – Black which any user can give it

rating range from 1 to 5 and all the total number of rating making it to be the most rated product followed by HP DeskJet 2130 All-in-One Printer- White which any user can give it rating range from 1 to 5 also and all the total number of rating making it to be the second most rated product until it reaches the least rated product.
From the table above, any product can be recommending to a user that have not seen that product before, by finding the correlation between the product we want to recommend and the product that correlate with it. And for this study, the product HP 2.4GHZ Wireless Optical Mouse and Sandisk 8GB MicroSDHC Memory Card will be selected, so that any product that correlate with them will be recommend to a user to buy.

Table 7. Correlation between the Sandisk 8GB MicroSDHC Memory Card and other correlated product

Title	Correlation	Number of Ratings
Sandisk 8GB MicroSDHC Memory Card	1.000000	51
Generic High Speed Transmission Micro SD 64GB	0.748353	26
Universal USB 2.0 Multi-function Card Reader	0.672556	23
Bluelans Keyboard Soft Case for Apple MacBook	0.536117	27
ASUS VivoBook X541SA- Intel Pentium Quad Core	0.377433	24
Seagate Multipurpose 500GB HDD	0.367538	25

Table 7 showed the correlation between the HP 2.4GHZ Wireless Optical Mouse and some other product that have correlation with it i.e. they are correlated. The selection of the correlation based on the product that have rating more than 20 ratings (ratings > 20) and after considering the rating more than 20 ratings, the first six products that have highest correlation with Sandisk 8GB MicroSDHC Memory was selected. And from this, any product from Generic High Speed Transmission Micro SD 64GB, Universal USB 2.0 Multi-function Card Reader, Bluelans Keyboard Soft Case for Apple MacBook, ASUS VivoBook X541SA- Intel Pentium Quad Core and Seagate Multipurpose 500GB HDD can be recommend to a user that have seen or buy the Sandisk 8GB MicroSDHC Memory but have not buy or seen the remaining five products.

Table 8. Correlation between the HP 2.4GHZ Wireless Optical Mouse and other correlated product

Title	Correlation	Number of Ratings
HP 2.4GHZ Wireless Optical Mouse	1.000000	91
Huiphoe Anti-Theft Business Laptop Backpack with USB	0.748353	22
Dell Inspiron 15 Intel Pentium	0.672556	25
Lenovo Ideapad 110-15ISK Gen Intel Core I3-6100U	0.536117	21
Universal Mouse Pad Red	0.377433	27
i-Life Zed Air Intel Celeron	0.367538	24

Table 8 showed the correlation between the HP 2.4GHZ Wireless Optical Mouse and some other product that have correlation with it i.e. they are correlated. The selection of the correlation based on the product that have rating more than 20 ratings (ratings > 20) and after considering the rating more than 20 ratings, the first six products that have highest correlation with HP 2.4GHZ Wireless Optical Mouse was selected. And from this, any product from Huiphoe Anti-Theft Business Laptop Backpack with USB, Dell Inspiron 15 Intel Pentium, Lenovo Ideapad 110-15ISK Gen Intel Core I3-6100U, Universal Mouse Pad Red and i-Life Zed Air Intel Celeron can be recommend to a user that have seen or buy the HP 2.4GHZ Wireless Optical Mouse but have not buy or seen the remaining five products.

4.4 Evaluation of Accuracy

The both matrix factorization method (Singular Vector Decomposition) and item-based method (Pearson Correlation) were evaluated and the MAE and RMSE from each value were obtained after conducting the evaluation. The lowest value for both MAE and RMSE were obtained while evaluating the matrix factorization method (Singular Vector Decomposition) and the average evaluation values for each metric are presented in table below:

Table 9. Evaluation of accuracy

Algorithm	RMSE	MAE
Matrix Factorization (SVD)	0.916250	0.708731
Item-based (Pearson Correlation)	0.930789	0.719434

5. CONCLUSION

Recommender systems are a powerful new trending technology for extracting additional value for a business from its user databases. These systems help users find items they want to buy from a business. Recommender systems benefit users by enabling them to find items they like, and help the business by generating more sales. Recommender systems are rapidly becoming a crucial tool in E-commerce on the Web. Recommender systems are being stressed by the huge volume of user data in existing corporate databases, and will be stressed even more by the increasing volume of user data available on the Web.
From the results of evaluation metrics, matrix factorization method (Singular Vector Decomposition) have RMSE of 0.916250 and MAE 0.708731 and item-based method (Pearson Correlation) have RMSE and MAE of 0.937089 and 0.719434 respectively. Which indicates that the matrix factorization method (Singular Vector Decomposition) scales slightly better than the item-based method (Pearson Correlation) when evaluating their prediction accuracy using RMSE and MAE. Recommendation system is still a new technology which has been around long time ago and which requires different expert from different field for its development. There are different techniques and algorithms that have been developed by different researchers. One among it is during the Netflix competition, which this online movies hosting company said anybody that can develop a good algorithms or methods for a recommender system will win $1million and different algorithms and techniques where developed but there is need to know which methods or techniques will outperform the others. Therefore, the following can be recommended for the further study: to study different recommender system methods or techniques e.g. content-based recommender system and check their accuracy so as to know which method is more accurate than the other.

REFERENCES

[Alb17] **Albert A. Y.** - *Matrix Factorization: A Simple Tutorial and Implementation in Python*, http://www.albertauyeung.com/post/python-matrix-factorization/ ,2017.

[A+99] **Aggarwal C. C., Wolf J. L., Wu K., Yu P. S.** - *Horting Hatches an Egg: A New Graph-theoretic Approach to Collaborative Filtering*. In Proceedings of the ACM KDD'99 Conference, San Diego, CA, pp. 201-212, 1999.

[BHK98] **Breese J. S., Heckerman D., Kadie C.** - *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52, 1998.

[C+08] **Chen L. S., Hsu F. H., Chen M. C., Hsu Y. C.** - *Developing recommender systems with the consideration of product profitability for sellers*. Int J Inform Sci ;178(4):1032-48, 2008.

[FLB11] **Francesco R., Lior R., Bracha S.** - *Introduction to Recommender Systems Handbook*, Chapter 1 in Recommender Systems Handbook, Springer New York Dordrecht Heidelberg London. ISBN 978-0387-85819-7, pp. 1-29, 2011.

[GU15] **Gorakala S. K., Usuelli M.** - Building a Recommendation System with R, Packt Publishing Limited, United Kingdom, ISBN 10: 1783554495, 2015.

[G+92] **Goldberg D., Nichols D., Oki B. M., Terry D.** - *Using collaborative filtering to weave an information tapestry*. Comm. ACM 35, 12, 61–70, 1992.

[K+97] **Konstan J., Miller B., Maltz D., Herlocker J., Gordon L., Riedl J.** - *GroupLens: Applying collaborative filtering to Usenet news*. Comm. ACM 40, 3, 77-87, 1997.

[R+94] **Resnick P., Iacovou N., Suchak M., Bergstrom P., Riedl J.** - *GroupLens: An open architecture for collaborative filtering of net news*. In Proceedings of CSCW, 1994.

[Vig13] **Vignesh P.** - *Big Analytics with R and Hadoop*, Packit Publishing, ISBNS:978-1-78216-328-2, 2013.