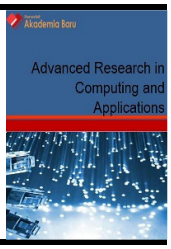
Penerbit Akademia Baru

Journal of Advanced Research in Computing and Applications

Journal homepage: www.akademiabaru.com/arca.html
ISSN: 2462-1927



E-Halal Restaurant Recommender System Using Collaborative Filtering Algorithm

Open Access

M. I. Mahadi[1], Nurulhuda Zainuddin[1,*], Norzatul Bazamah Azman Shah[1], Nur Asyira Naziron[1], Siti Fatimah Mohd Rum[1]

[1] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Jasin Campus, 77300 Merlimau, Melaka, Malaysia

ARTICLE INFO

Article history:
Received 5 March 2018
Received in revised form 12 April 2018
Accepted 10 July 2018
Available online 4 August 2018

Keywords:
Collaborative filtering, K-means clustering, recommender system, multi-criteria rating system

ABSTRACT

In this paper, a web-based recommender system is proposed for Muslim consumers to select their favourites Halal food restaurant based on user ratings. The user will rate several restaurants according to food, price, decoration and service which they have visited. The recommendation of the restaurants will be calculated using collaborative filtering technique. Collaborative filtering technique is one of many techniques used in recommendation system. İt has been applied in many applications. Collaborative filtering technique uses user's preferences and their neighbours to give recommendation. In the collaborative filtering, it uses K means clustering algorithm to calculate the similarity between users. Similar users will be clustered together and recommended the same restaurants. Finally, the application in the future could make Halal restaurants rising in demands and promote tourism in Melaka.

1. Introduction

1.1 Halal Background in Malaysia

Malaysia has been among the frontier in Halal industry. Its pioneering in the industry has begun in 1974 when the Research Centre for the Islamic Affairs Division in the Prime Minister's Office started to issue Halal certification letters for products that met the Halal criteria. The first standard that was established in 2000 has become an important milestone for Malaysia in venturing the Halal industry. It is estimated that the Halal industry constituted USD2.30 trillion of global market value [1].

High awareness among Muslim consumers regarding Halal food also has contributed significant demand towards Halal food. In a research by Ambali and Bakar [2] towards respondents in Shah Alam, more than 70 percent of the respondents are conscious about the status of their Halal food. Exposure, religious belief, health related reasons and the Halal certification plays important role to

* *Corresponding author.*
E-mail address: Nurulhuda Zainuddin (nurulhuda066@melaka.uitm.edu.my)

create awareness among Muslim consumers. The mass media also play important role to educate the Muslim community about Halal concept. Having Halal food does not mean they obey to their religious belief solely, they also believe that food with Halal certification also guaranteed with its cleanliness and healthy.

Halal certification in Malaysia is conducted by JAKIM. The process of getting Halal certification is very strict and require many stages to ensure the status of halal in Malaysia is not abused. Collaboration with Department of Standards Malaysia, Institute of Islamic Understanding Malaysia and Malaysian Institute of Research and Standard (SIRIM) has enabled the halal brand to go further by standardising it using MS1500:2009: Halal Food –Production, Preparation, Handling and Storage –General Guidelines (Second Division) which is to standardise Halal requirement of production, preparation, handling, and storage of Halal food and beverages [7].

In this paper, a web-based recommender system is proposed for Muslim consumers to select their favourites Halal food restaurant based on user ratings. A recommender system of the Halal restaurant web-based application will be developed which is based on four criteria which are food, service, decoration and price. The system then will be tested to test the system's recommendation accuracy. According to Symoneaux *et al*., [5] knowing how much consumers like a product is not enough, it is important to understand the reason of their choices. From the statement, it is important to understand user's preferences. A multi criteria rating system is used to obtain and understand user's preferences of what they like and dislike. Besides, according to a research by Gerasimou and Papi [6], a large empirical work has suggested that the existence of "choice overload" could affect consumer decision making ability. People tend to avoid or indefinitely defer choice when faced with large menus of alternatives. Therefore, recommendation system is here to help people to make a better choice [4].

2. Related Works

The detail discussion about rating system in recommender system and the application, will be explained further in this section. Methodology of recommender system and existing applications in recommender system also will be discussed to see the overall view of the proposed system.

2.1 Personalization Technology and Recommendation System

Having too much of choices is unhealthy according to a research by Reed *et al*., [3]. Decision making of an extensive array of choices could lead to aversion, often leading negative emotional states and poor behavioral implications. This negative effect of too much choices could affect the efficiency of decision making and is called as "paradox of choices" [3]. Therefore, according to Adomavicius and Kwon [4], personalization technology and recommender system are invented to overcome this problem by providing personalized suggestion regarding which information is the most relevant to be chosen by the user. Implementation of this system could make revolutionary in decision making. It also could bring more market value in certain products.

Multi-criteria rating system could leverage a recommender system in many ways. It could improve the accuracy of providing suggestions to users. According to Adomavicius and Kwon [4], multi-criteria ratings can be successfully leveraged to improve recommendation accuracy compared to single-criteria recommendation techniques. One way to get user feedback is through rating system. A rating system usually will provide a range of score for user to fill. The score will determine user's satisfaction towards particular item. Many applications nowadays use rating system in order to obtain user's feedback. Its simplicity, convenience and effectiveness to improve customer service and business operation has led many studies to improve its efficiency. Multi criteria rating system

has been introduced to gather more details from customer feedback. Its popularity has increased in many industries. Online shopping websites like Bestbuy.com and Amazon.com use multi criteria rating system to acquire feedback from their customer [4]. Compare to single criterion rating system, multi criteria rating system can provide additional information about user's preferences. This can be beneficial to recommender system since the additional information can improve accuracy to the recommendation system.

Recommender system and personalization technology are one of technology contributions in web applications. Their existence is to assist human to make a better choice and decision. This is because the internet can provide a lot of information, however overload information can cause inefficient decision making. Recommender system is used to provide suggestion to user regarding which information is the most relevant. Therefore, accuracy and efficiency could be improved in decision making. Nowadays, many online shopping and other applications use recommender systems in their system. Among the popular sites are Amazon.com, eBay, Netflix, Lazada and Trivago [4]. The objective of recommender system is to predict for the yet unseen items based on user's feedback and activities. The information is collected when users offer their feedback on purchased or consumed items. The recommender system will then recommend or suggest items with the most precise estimated relevance to the users [4]. There are five type of recommendation techniques; collaborative filtering-based recommendation, content-based recommendation, demographic recommendation, knowledge-based recommendation and hybrid recommendation.

2.1.1 Collaborative filtering technique

Most commonly used technique is collaborative filtering. The technique will try to automate the process by aggregating user ratings on items. Then, it will find similar users based on their ratings and finally recommending items based on user similarities. Compared to other technique, collaborative filtering does not contemplate with the item's content. This technique particularly useful in rendering complex items such as movies, videos, music and etc [8].

To test how the technique works, given user *u* and item *i* that has not been rated yet by user *u.* Collaborative filtering technique will try to predict rating on item *i* for user *u*. The recommendation process will be constructed in two steps. First of all, a set of user's *u* having similar preference will be selected on the basis of previously rated items. Second, items *i* that has been rated by user *u* will be accumulated [10]. A speculative approach consists of considering the items similar to *i* rather than considering users of *u* and accumulating rating selected by *u* to those items. There are two most important type of collaborative filtering algorithms. They are memory based and model based.

The memory based collaborative filtering is a method that is based on the user and item rating matrix. There are two steps in memory based. In a given set, there are user *u* and item *i*. In the first step, a set of U_u^k will be selected based on *k* and its closest neighbour of *u*. The distance between users can be establish by defining an appropriate user similarity function. The second step is approximating the rating of item *i* on the set of U_u^k. There is a function where an approximation of the rating can be calculated. The rating can be calculated using the mean arithmetic rate or using weighted majority voting over U_u^k. In a paper by Garcin *et al*., [10], it compares three different function in aggregation of rating system; mean, median and mode. According to him, the median function result improves the recommendation accuracy and robustness. Nevertheless, the main issue in the algorithm is the sparsity of user-item matrix. It means that in large scale system, users will only rate small fraction of available items. Hence, the similarity between a given two users will be small. This issue could result in poor prediction of user rating [10].

Meanwhile, model based collaborative filtering build a predictive model based on historical data. Instead of using user-item rating matrix, this algorithm uses the historical user-item information during the learning phase. It then will create learned model and make predictions on the test data. Among the proposed examples of model based collaborative filtering are Bayesian model collaborative filtering, clustering models collaborative filtering, and dependency network collaborative filtering [10]. Table 1 shows the comparison of techniques used in collaborative filtering-based recommendation.

Table 1
Comparison of techniques used in Collaborative Filtering-Based Recommendations

Author	Technique	Domain	Result
Ma, Yang, Wang, Li, and Li, 2014 [11]	Self-organizing mapping (SOM) based improved k-means clustering	TV recommendation system	The median absolute eror of the proposed technique was rediced by 15.7% and 17.4% compared to improved k-means and the k-means algorithm. The proposed algorithm can efficiently improve the TV recommendation system.
Kwon, 2011 [12]	Heuristic-Based Ranking Approaches to Improve Aggregate Recommendation Diversity	Yahoo!Movies dataset	The results show that up to 20-25% diversity, it only loss 0.1% accuracy, and when up to 60-80% diversity gain, only 1% accuracy loss, and even substantially higher diversity improvements up to 250% if some users are willing to tolerate higher accuracy loss
Hu, 2014 [13]	Indifference relation	Group buying website	From the test result using analysis of variance (NOVA), the average result of the precision in top 1, precision in top 3 and precision in top 5 metric are equal to each other, and rejected at 5% level. The turkey test also shows that the precision top 1 and 3 outperforms the precision top 5 measure. In conclusion, the proposed recommendation approach has better performance when few most highly ranked initiators need to be provided to users.
Cho, Moon, Noh, and Ryu, 2012 [14]	K-means clustering of item category based on Recency, Frequency and Monetary (RFM).	e-shopping mall	Compared to traditional collaborative filtering algorithm, the proposed algorithm is higher 18.3% in precision, 14.03% in recall, and 9.23% in F-measure. As a result, this system could be used immediately in u-commerce under ubiquituous computing environment which is required by real time accessibility and agility of fnishing particular tasks.

2.1.2 Content-based recommendation technique

Another type of recommendation techniques is content-based recommendations where it is purposely to identify interesting Websites or articles and recommending them based on their content. This technique will process the content of the items and recommend them based on user's profile. The user's profile is constructed based on content of items that the users previously liked or disliked. The items will be represented via a set of relevant attributes extracted from their content is far from being a trivial task. And it is strongly depending on the kind of items at hand [8]. Vector Space Model is an effective way to represent the content of items where an item is represented as a weighted vector of features.

In contrast, the demographic recommender is used in categorizing users based on personal demographic profile. It will then recommend items or products based on the user's demographic profile. The user's demographic profile can be collected via personal information like the age, address, language and etc. Through the recommendation system, the system will then make conclusion that users with the same profile will have similar interest. Nevertheless, collecting demographic data can be difficult due to data privacy and security regulations. However, this system can solve the cold start recommendation problem which does not need any information of user's history purchased or visit [10]. In the paper by Weber and Castillo [15] show that how demographic data can be used to offer recommendations. The main objective of their paper is to identify how different demographic data will affect user's Web search behaviour. They use three sources of information. The sources are a subset of Yahoo! U.S. query log, profile information provided by the user and U.S.-census information. The result is different users with different demographic features like age and income will require different personalized recommendations. Thus, this bring advantages of having to train less data, protect user privacy and a simpler interpretation.

2.1.3 Knowledge-based recommendation technique

Besides, knowledge-based recommendation system is based on rules, patterns and functional knowledge. Functional knowledge is an information of how a given item can meet a specific user in the particular application domain. Knowledge based recommendation system can retrieve data of categories of items from e-commerce web site and generate suggestion that will meet user's preferences. The knowledge-based recommendation system however has two drawbacks. Its ability to suggestion might be static and the pruning techniques error in the information extraction process. Nevertheless, the availability of domain knowledge can alleviate the effect of cold start problem [10]. In hybrid methods, combination of multiple technique is implemented to satisfy limitations of single recommendation technique. In the book of mining user generated content by Garcin *et al*., [10], he proposed a taxonomy for hybrid technique recommendation classifying them into seven categories. The categories are weighted, mixed, switching, feature combination, feature augmentation, cascade and metalevel.

The most suitable technique to implement in our proposed system will be collaborative filtering technique due to its simplicity and easy to use. This technique also is the most popular technique used in many applications. In recent applications, the technique has been improved with many methods. The accuracy also has been improved. The technique also combined with many techniques to further improve its accuracy and efficiency.

3. Methodology

Furthermore, the methods of how the recommender and multi-criteria rating system will solve data overloads are identified from research materials. The system is the core of the project as it will act as engine and drive the system. Details of the recommender and multi-criteria rating system has been discussed further in literature review. In the literature review, methods and algorithms used and implemented in most recommender system is discussed and compared before we proposed the suitable algorithm.

3.1 Designing System Algorithm

From the research and study conducted, collaborative filtering is chosen as the method how the system will be created. The collaborative filtering algorithm is carefully analysed and considered to be implemented in the system. K-means clustering method is used as the algorithm technique.

$$d\left(user_i, user_j\right) = \sqrt[\gamma]{\sum_{k=1}^{n}\left(ri_k - rj_k\right)^{\gamma}}$$

Fig. 1. Euclidian Distance [9]

Clustering method could be defined as the process of organizing objects in a database in to clusters or groups. The objects which belong to the same group can be concluded as having high degree of similarity while different group could mean the objects have high degree of dissimilarity. Clustering method falls into undirected mining tools. The goal is to discover structure in the data as a whole. In order for objects to be grouped together, two main criteria must be satisfied; each group or cluster is homogenous. In other words, the objects which belong to the same group must be similar to each other. On the other hand, if the objects are belong to different group or cluster, they must be different to each other [13]. K-means clustering is a non-supervised learning algorithm. The algorithm uses as input a predefined number of clusters, *k* that is derived from its name as initial cluster centres and calculate the distance between each object as initial cluster centre and calculate the distance between each object and each cluster centre. Next, it will put each object to the nearest cluster and update the averages of all clusters. The process will repeat until the criterion function converged [11].

In this approach, theoretically the system will act as an intermediary to give suggestions to user. First of all, the user will need to give rating scores based on four criteria which are food, price, services and decoration restaurants they like. Based on the rating score, the system will group the user with other users which have similar preferences with the current user. The rating score is calculate using Euclidian distance as shown in Figure 1 which to predict which cluster the user will belong to. After the cluster for the current user has been determined, the system will give suggestions to the current user which the user never visited before based on other user's preferences in the same cluster.

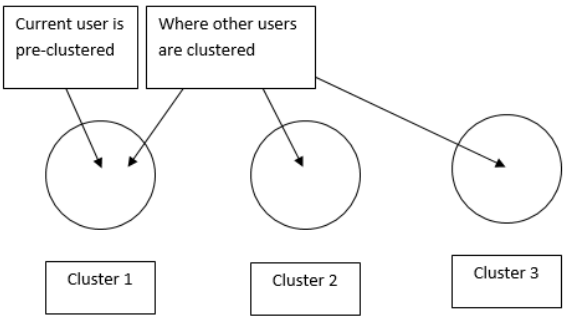


Fig. 2. User is Pre-clustered

From above Figure 2, it visualizes how the user is being clustered and suggestions is given. The current user is pre-clustered based on the review given. From the rating score let say the user put 1 as total rating score, the user will be pre-clustered in cluster number one. Then the system will give suggestions to the current user based on the other user which have similarity with the current user. Whichever restaurants that current user has not rated yet but other users have, that restaurants will be suggested to the current user. Figure 3 show the proposed framework used in the development.

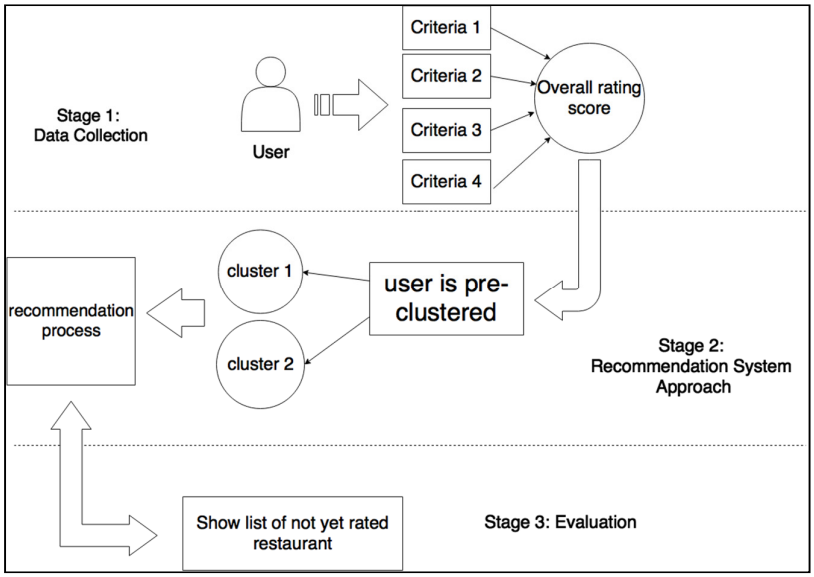


Fig. 3. Proposed framework

Meanwhile the multi-criteria rating score will be aggregated using mean arithmetic. In a paper by Garcin *et al*., [10], there are three criteria to be considered when aggregating multiple rating score. They are based on informativeness, robustness, and strategy proofness.

$$pred_{u,i} = \overline{r}_u + \frac{\sum_{v \in V} sim_{u,v}(r_{v,i} - \overline{r}_v)}{\sum_{v \in V} |sim_{u,v}|},$$

Fig. 4. Mean Formula [10]

Mean is chosen to calculate the total rating score by user which is shown in Figure 4. The mean score will then have stored as overall score of the restaurant rated by the user which shown in Figure 5. The score then will be used to calculate similarity index in the recommendation system algorithm. The rating of each of criterion can be represented using star, icon or number.

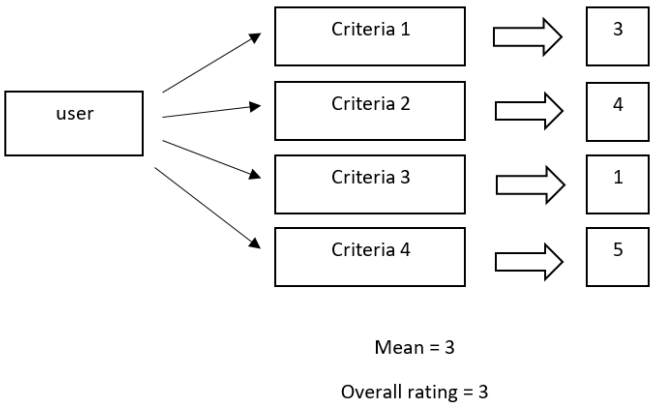


Fig. 5. Mean rating score is selected to be the overall rating

3.2 Project Development

Figure 6 shows the homepage of the system for registered or unregistered users. Users can see the list of halal restaurants which has been certified by JAKIM in Melaka. This page also shows the latest reviews and ratings given by users that have visited the halal restaurants in Melaka.

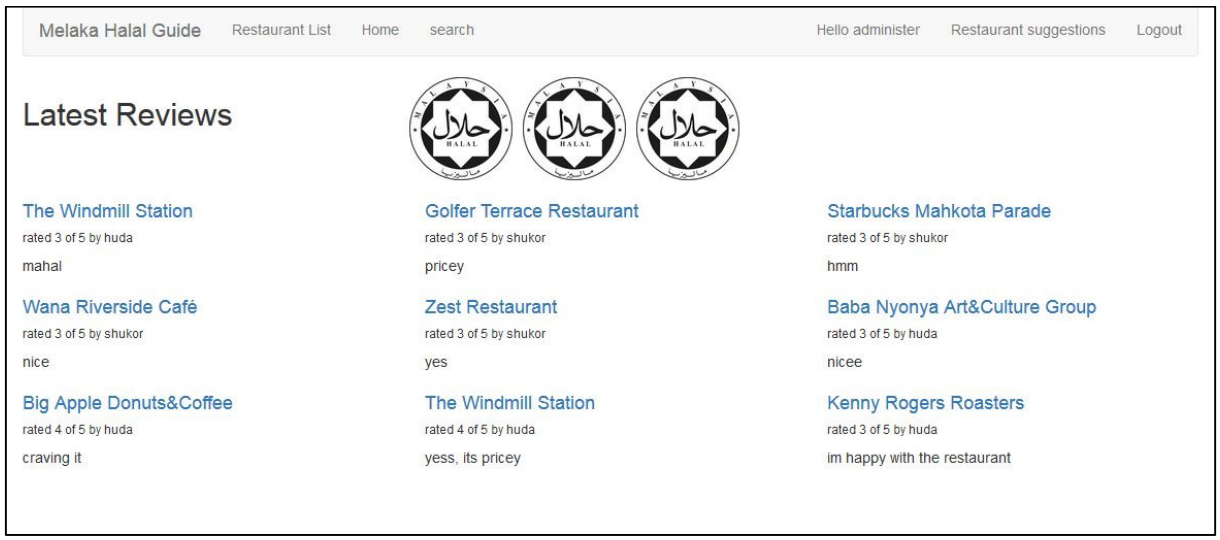


Fig. 6. Latest review by users

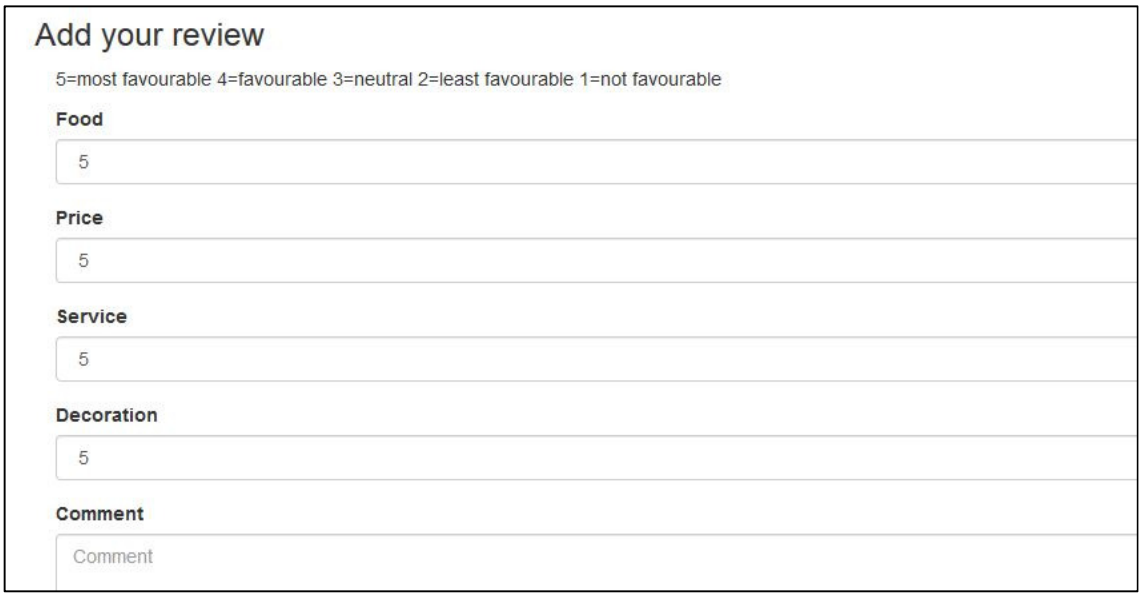


Fig. 7. Criteria users will rate

Figure 7 is the form which users will fill accordingly. There are four criteria which we surveyed would be sufficient to rate a restaurant. In the form, we put a highest rate of five as the most favourable till one as the most unfavourable to user as a guide to give rating score. From the four ratings given by user, a total average will be calculated and inferred as the total score by user.

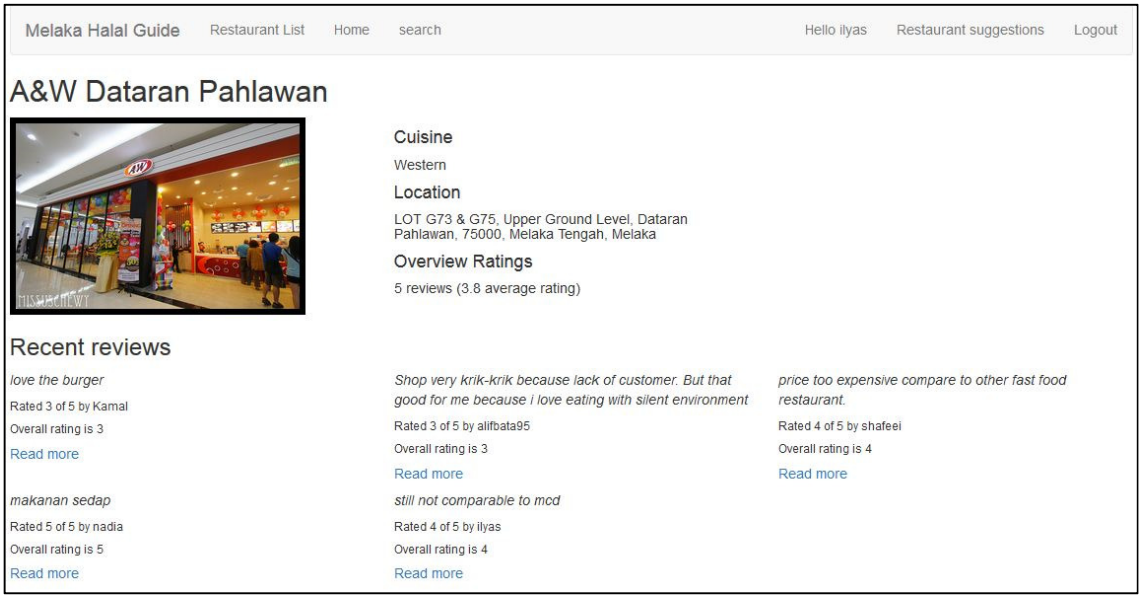


Fig. 8. Recent reviews by users

Whenever user give rating, the rating information will be populated here. All user's reviews will be recorded here at the restaurants page as shown in Figure 8. This page also provides a brief

information about the restaurants. This page will give a guidelines and information to new users about the restaurants they are going to visit.

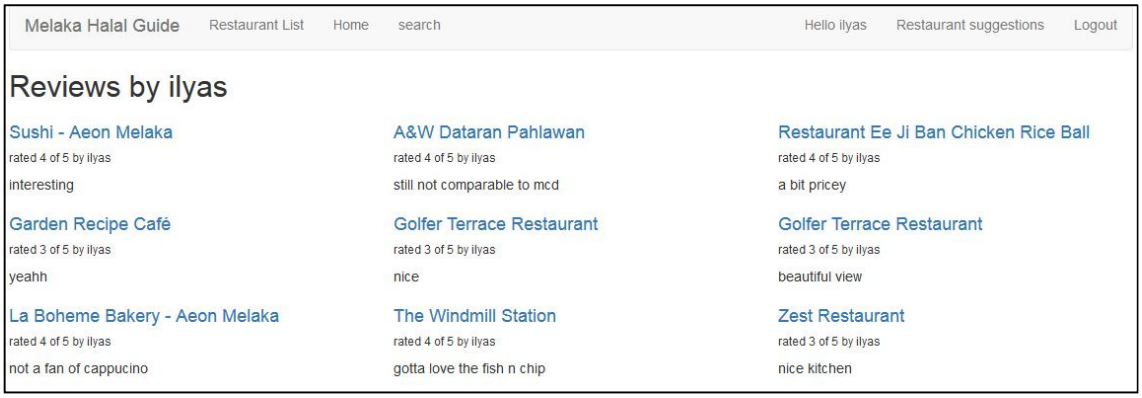


Fig. 9. Reviews by active user

Figure 9 shows all reviews by the current active user that will be populated. This page can be as guidelines by the user to view which restaurants they have visited and gave rating. A total rating given by user will be calculated by the system and produced a recommendation to the user in the recommendation page.

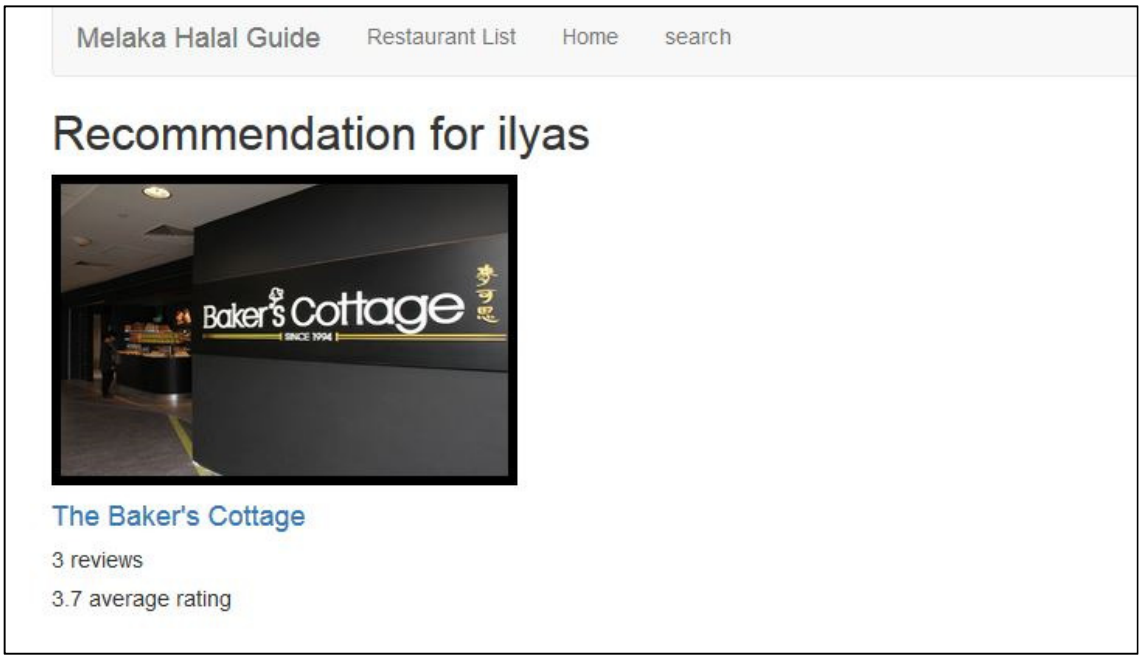


Fig. 10. Recommendation page to user

Figure 10 shows the page where recommendation results are displayed to user. From restaurants that has been reviewed by user in Figure 9, a total score will be calculated and the user is assigned in a cluster. Restaurants that will be recommended to the user is based on other users which belong to same cluster. The system will assume that users which belong to the same cluster has same

preferences and will suggest restaurants which the other users have visited but never been visited by the active current user. The system will predict that the current active user will like the restaurant too.

4. Results and Discussion

4.1 Testing Background

The system is built using Django framework to implement the collaborative-filtering method. Tools and plug ins provided in Python make it easier to build the recommender system. yThe functionality testing is conducted by several UiTM students who have visited restaurants around Melaka. There are 10 volunteer respondents who willingly take the test and give opinion regarding the accuracy of the system in giving recommendation. The test comprises of six male students and four female students. From the survey, four respondents have visited around five restaurants in Melaka and four of the respondents have visited around 10 restaurants around Melaka. Meanwhile, one respondent has visited around 15 different restaurants in Melaka and the remaining one respondent has visited 20 different restaurants in Melaka.

4.2 Testing Methodology

The system testing is to obtain the respondent's review about the accuracy of the system in giving suggestion to them. First of all, the respondents need to register an account in the system. Then, they need to give review at least four to six recommendations in order to avoid the 'cold start' problem. After that, they will see the list of restaurants which the system has suggested for them. From the list, they need to fill a survey to find out their review about the system accuracy in providing recommendation.

4.3 Testing Results

From the Table 2, all of them agree that food, price, service and decorations are suitable criteria to rate a restaurant. Hence, it is safe to say that the criteria are sufficient to rate a restaurant.

Table 2
System Testing Results

	Percentage Respondents Agree	Accuracy
Respondent's review regarding the system's accuracy in giving suggestion.	50%	51-75%
	40%	75-100%
	10%	25-50%
Respondents agree that website is useful as a guidance to tourist and Muslim consumers to find halal certified restaurants.	-	80%
Respondents will use the website to find halal certified restaurants around Melaka.	-	90%

From Table 2, half of the respondents agree that the system can give 51 percent to 75 percent accuracy in giving suggestions and 40 percent of the respondents agree that the system can give 75 percent to 100 percent accuracy in giving suggestions. Meanwhile, only 10 percent of the respondents agree that the system give 25 percent to 50 percent accuracy in giving suggestions.

Besides 80 percent of the respondents agree that website is useful to tourist and Muslim consumers to find halal certified restaurants. Meanwhile, 20 percent of the respondents feel unsure. Then, 90 percent of the respondents will use the website to find halal certified restaurants around Melaka. Meanwhile, only 10 percent unsure if they want to use the website in the future.

5. Conclusion

5.1 System Limitation

This system inherits the 'cold start' problem of collaborative filtering. Means that the system cannot gives recommendation if the user has not rated any restaurants yet. Besides, the user also needs to at least rate approximately four to six restaurants first before the system can give recommendation to the user. Moreover, this system requires more review data to give more accurate results of giving recommendation to user. If the system lacks review data, possibly the system will provide less accurate suggestion. Last but not least, data obtained from Jakim's Halal directory is insufficient. A lot of more popular halal restaurants in Melaka have not applied for halal certificate since it is not compulsory to apply for halal certificate from Jakim. Their awareness towards halal certification still low compared to established restaurants like A&W and Starbucks.

5.2 Future Works

This system could become a pioneer to a better recommendation system. The accuracy can be improved by applying other algorithms or combine with other algorithms such as hybrid algorithm to overcome the 'cold start' problem. Besides, recommendation also could be given by using input from user's activities such as frequency of user visits the same restaurants, type of restaurants or their food of choices. However, using those input may require user's consent as to avoid data privacy breach policy.

Acknowledgement

The authors wish to thank Universiti Teknologi MARA Cawangan Melaka for the administrative support during the completion of the research.

References

[1] Islamic Tourism of Malaysia. MALAYSIA – THE WORLD'S LEADING HALAL HUB [Internet]. 2017.
[2] Ambali, Abdul Raufu, and Ahmad Naqiyuddin Bakar. "People's awareness on halal foods and products: potential issues for policy-makers." *Procedia-Social and Behavioral Sciences* 121 (2014): 3-25.
[3] Reed, Derek D., Brent A. Kaplan, and Adam T. Brewer. "Discounting the freedom to choose: Implications for the paradox of choice." *Behavioural processes* 90, no. 3 (2012): 424-427.
[4] Adomavicius, Gediminas, and YoungOk Kwon. "New recommendation techniques for multicriteria rating systems." *IEEE Intelligent Systems* 22, no. 3 (2007).
[5] Symoneaux, R., M. V. Galmarini, and E. Mehinagic. "Comment analysis of consumer's likes and dislikes as an alternative tool to preference mapping. A case study on apples." *Food Quality and Preference* 24, no. 1 (2012): 59-66.
[6] Gerasimou, Georgios, and Mauro Papi. "Duopolistic competition with choice-overloaded consumers." *European Economic Review* 101 (2018): 330-353.
[7] Ibrahim, S., and M. Othman. "Developing and Validating Halal Service Quality Instrument for Malaysian Food Service Establishments: A Conceptual Paper." *Procedia-Social and Behavioral Sciences* 130 (2014): 400-408.
[8] Moens, Marie-Francine, Juanzi Li, and Tat-Seng Chua, eds. *Mining user generated content*. CRC Press, 2014.
[9] Dakhel, Gilda Moradi, and Mehregan Mahdavi. "A new collaborative filtering algorithm using K-means clustering and neighbors' voting." In *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on*, pp. 179-184. IEEE, 2011.

[10] Garcin, Florent, Boi Faltings, Radu Jurca, and Nadine Joswig. "Rating aggregation in collaborative filtering systems." In *Proceedings of the third ACM conference on Recommender systems*, pp. 349-352. ACM, 2009.

[11] Ma, Zhaocai, Yi Yang, Fei Wang, Caihong Li, and Lian Li. "The SOM Based Improved K-Means Clustering Collaborative Filtering Algorithm in TV Recommendation System." In *Advanced Cloud and Big Data (CBD), 2014 Second International Conference on*, pp. 288-295. IEEE, 2014.

[12] Kwon, YoungOk. *Computational Techniques For More Accurate and Diverse Recommendations*. University of Minnesota, 2011.

[13] Hu, Yi-Chung. "A multicriteria collaborative filtering approach using the indifference relation and its application to initiator recommendation for group-buying." *Applied Artificial Intelligence* 28, no. 10 (2014): 992-1008.

[14] Cho, Young Sung, Song Chul Moon, Si Choon Noh, and Keun Ho Ryu. "Implementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM." In *Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on*, pp. 378-383. IEEE, 2012.

[15] Weber, I., Castillo, C., & '10), P. of the 33rd international A. C. M. S. conference / R. and development in information retrieval (SIGIR '10) P. of the 33rd international A. C. M. S. conference / R. and development in information retrieval (SIGIR. (2010). The demographics of web search TT -. In Annual ACM Conference on Research and Development in Information Retrieval TA - (pp. 523–530). ACM, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, USA.