



REVIEW BASED RECOMMENDER SYSTEM FOR BUSINESS ADVANTAGE USING NATURAL LANGUAGE PROCESSING

Furkan Ahmad* and Akheela Khanum M

Department of Computer Science and Engineering, Integral University Lucknow, India

ARTICLE INFO

Article History:

Received 6th February, 2018

Received in revised form 20th

March, 2018 Accepted 8th April, 2018

Published online 28th May, 2018

Key words:

Recommender systems; Review Based Recommender System; Hotel Recommendations; Natural Language Processing; Hotel-Review.

ABSTRACT

In this current age where everyone complains about time management issues, the extent to which an individual depends on others recommendation about a particular entity is enormous. Recently, the state-of-the-art recommender systems are on surge which basically intends to predict rating or preference given by a user to an item. This paper presents a Review Based Recommender System for Hotel Recommendations which recommends hotels in particular cities, which enables users to identify the most suitable hotel present in those cities based on users preferences and other users rating about the hotels. The rating of a hotel from previous users is determined by applying Natural Language Processing approach on a Hotel Review Compilation and a hotel-review database is created. This technique can be utilized for business advantage and it can help users identify the best hotels in a particular city in an uncomplicated and informal manner quite easily.

Copyright©2018 Furkan Ahmad and Akheela Khanum M. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In this fast moving world where every individual needs the best of services, in a blink of an eye, there should be some system that helps the users in choosing the best amongst the available options. In order to arrive at a conclusion about choosing any service, one has to have adequate information about the services. For the past decade or so, there has been rapid development in the field of technology which has enabled millions of user's across the globe to easily access any information about any service, whether it be in the field of education, hospitality, travel, gaming, electronics and in almost every field. Lately, the hospitality and travel industry has grown exponentially. The rise in these industries has enabled the user's to choose from a bunch, in order to get the preferred one. There are many factors which affect user's choice, namely price, location, quality, food to name a few. Although there exists a surplus amount of hotels available, to find one according to your choice remains a quite daunting task.

In the past decade, a lot of research has been put into automating the process of user's rating and reviews. Usually, the users search for hotels which have impressive "star rating", this approach has its limitation in a way as it does not examine the overall facilities of the hotel. Recently, as part of service improvement, the hotels lay greater emphasis on reviews from

the users. This has made the hotel industry much more competitive and has given an edge to the users in the way of selecting suitable hotels.

The use of recommender systems comes to the rescue for providing these facilities to the users. Basically, recommender systems are applications which are used by e-commerce websites to recommend products and services to the users in a personalized manner, according to their preferences as described in [1]. Recommender systems are of mainly two types namely: (1) Collaborative Filtering Systems (2) Content-Based Systems. Collaborative filtering systems recommend according to the reviews given by a particular user's neighborhood i.e the ones who have similar taste in particular services might have interest in another similar service. Content-Based systems compare a particular user's profile with the content of the items available. Recommender systems filter useful information as per the user requirements and interests.

These recommender systems have problems in dealing with these reviews collected as these reviews are in the natural language. The reviews need to be processed efficiently in order to derive some meaning whatsoever from the review. Such a technique is discussed in [7] which proposes a Keyword Aware Service Recommendation (KASR), in which keywords are exploited to indicate users preferences and the candidates quality of services. Another issue which needs to be addressed by the recommender systems is the diversity of the users available on the internet. For example, let's say if a single

*Corresponding author: **Furkan Ahmad**

Department of Computer Science and Engineering, Integral University Lucknow, India

person is looking for a hotel his/her preferences would be low cost, cleanliness can be compromised up to an extent which is in stark contrast to a family looking for a hotel in the same city, the preferences will comprise of safety, cleanliness, good connectivity, healthy food easily available and many more. The recommender systems should come closer to the human level recommendation which is devoid of such drawbacks. The amount of data available in hotel industry i.e reviews are not so abundant as compared to the reviews in e-commerce websites, as the amount of purchaser of products online are humungous in quantity whereas travelers are significantly lower. This paper aims to recommend hotels to users in major cities around the globe in a simplified manner. This study has its focus on collecting the reviews and cleaning the reviews as they are in natural language i.e they contain certain words which are limited to the social media only such as gr8,awsm,omg, and the list goes on. To improve the accuracy of the proposed recommender system these words are processed from the reviews.

Further, the paper is distributed as follows: Section 2 covers brief literature survey whereas Section 3 discusses our proposed offline system following with the results achieved in Section 4. Finally, Section 5 concludes the discussion along with some future scope.

LITERATURE REVIEW

Over the years, Recommender systems have evolved significantly to emerge as a solution for online users. This paper [10] suggests a review based service recommendation which suggests services to the users. This study [3] explores the various characteristics and different potentials of numerous prediction techniques in recommendation systems.

The technique used in [4] is AWSR (Active Web Service Recommender) which is predominantly based on user's usage web history and ultimately suggests web services to users, this technique shows excellent performance while doing the above mentioned task. This paper [13] further suggests a user-based collaborative filtering (CF) algorithm on a cloud platform. The proposed technique in [11] is a novel approach which generates latent embeddings for podcast items by utilizing all the information available about the audio-items. This study [12] proposes the unification of Collaborative filtering and Content-Based recommendation as these two have been the main focus of researchers lately. This research [8] further extends the scope of recommender systems to recommend sports videos which are available on the internet, this study mainly focusses on audio-visual content.

This study proposes a technique for music recommendation [6], which learns from a sample of collaborative filter data and further uses it for optimizing content-based similarity. Recommender systems can also be used to enhance security as in [5], which studies posts on the social platform to study whose posts are positive and whose negative, subsequently monitoring the possible impacts of negative posts on the nation or a society. A novel technique was discussed in [7], introducing KASR (Keyword Aware Service Recommendation) method which recommends a personalized service recommendation list.

An interesting technique is proposed by [1], which tracks active tourists and predicts their interests, which are usually based on the collection of other tourists preferences and taste

information. Recently, an approach discussed by [9] shifts from traditional recommender systems which uses a single criterion to recommend items, by introducing multiple criterions for each item, which has high potential to outperform the traditional methods, in the near future. Although there are abundant recommender systems available, there remains a need for feedback on other users review. This study [2] proposes a new method based on collaborative filtering which takes into account other users feedback about the reviews available.

Proposed Work

Offline Review Extraction

The human recommendation is a clear winner against these recommender systems, but what needs to be looked upon here is the enormous amount of feed a human brain receives and the brain easily filters relevant information from this feed/data. A recommender system does not possess this luxury, on the contrary, the data and source knowledge remains the heart of these systems.

The hotel, tourism, jobs recommendation systems are most complex recommendation systems to build due to the availability of data in the unstructured form. Any recommender system would require data in a structured form. This section discusses the extraction of data and further mining of the text found in the review is done in order to make it suitable for information retrieval and finally calculate the polarity of the reviews and finally analyzing the sentiment of the writer whether it be positive, negative or neutral. In this paper we have extracted the dataset on hotel reviews from UCI: Machine Learning Repository which contains reviews of hotels in 10 cities namely (Dubai, Beijing, London, New York City, New Delhi, San Francisco, Shanghai, Montreal, Las Vegas, Chicago) containing 80-700 hotels across each city and approximately 259000 reviews are used to build the proposed recommender system. However, the data extracted is in dire need of some form of structuring and we have carried out following steps on the dataset to extract meaningful knowledge from the dataset. The steps are as follows:

Database Creation: As we know that the data extracted from any source is initially extracted into an XML file. These files are then stacked to the database.

Review Cleaning: The reviews in the database contain discrepancies which need to be addressed. The following steps are taken to address these reviews:

1. **Same Review Removal:** The reviews which match completely with a review already present in the database are ignored.
2. **Removal of all Non-ASCII characters:** Every text file present on the internet or anywhere is in ASCII (American Standard Code for Information Interchange).The Characters which are not in ASCII format are stripped.
3. **Text Normalization:** Text Normalization is done for converting text to speech. Numbers, abbreviations are bound to be pronounced differently, as per the context of the sentence. For example, vii can be pronounced viie, viiie or even the numeric seventh. The text needs to be present in the normalized form.

4. **URL Removal:** The URL (Uniform Resource Locator) does not translate into any suitable information, thus, needs to be excluded from the reviews.

Review Language Correction: There are several words which are not part of the standard dictionary but they are predominantly used in our daily conversations. For example. Wth! The service sux will be considered as a neutral review due to the non-existence of these words in the dictionary. On the other hand What The Hell! The service sucks will be taken as a negative review. To counter this issue we have abbreviated few words which are constantly found in these reviews, a library is created for these words. Whenever any of these words are found it gets replaced with the appropriate abbreviations present in the library.

After all the above preprocessing steps, the dataset is ready for some useful information retrieval. The arrangement of the dataset after these steps is illustrated in figure 1 given below.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	City	Hotel Name	Review Title	Review										
2	Dubai	ABC Almanar Hotel Apartment	Good overall service	Just came back after a week at this hotel. The Al Manar hotel apartments (part of ABC group) is located a 10 min										
3	Dubai	ABC Almanar Hotel Apartment	OK rooms construction next door	Room was nice and modern. Had reasonable size bedroom - large tv, kitchen, big fridge so, of for extended stay. I										
4	Dubai	Admiral Plaza Hotel	Value for money	The hotel has very good location - 10 minutes walking distance to Dubai Museum and the creek, easy to find. It										
5	Dubai	Admiral Plaza Hotel	Good Family Hotel	It was a good experience as the Hotel was situated at the heart of Dubai and from there everything was near th										
6	Dubai	Admiral Plaza Hotel	Good hotel	Good hotel offering value for money. Breakfast was simple but sufficient. Hotel is in a good location. Al Ghubaib										
7	Dubai	Admiral Plaza Hotel	Not the Burj Al Arab but a safe centrally loca	I am a single female traveler who stopped over in Dubai for three nights on my way to Rwanda. Being alone my										
8	Dubai	Admiral Plaza Hotel	Very Affordable	We were a group of 5 adults and 2 kids. The hotel is very affordable and it is value for money. The restaurant fo										
9	Dubai	Admiral Plaza Hotel	efficient service	Stayed at this hotel in may/ june/2008 Ramp; found the property conveniently located in Bur Dubai.Found the										
10	Dubai	Admiral Plaza Hotel	Value for Money	I had stayed in Admiral Plaza Hotel for business purpose. As my area of interest is Bur Dubai and Deira, the hot										
11	Dubai	Admiral Plaza Hotel	Great experience	Having read previous travellers reviews... I was apprehensive on arrival. But I can honestly say that my experie										
12	Dubai	Admiral Plaza Hotel	The facilities advertised dont work	While the rooms are clean, some of the services provided and advertised are not up to the mark. The Wi Fi refu										
13	Dubai	Admiral Plaza Hotel	great position and staff	Admiral Plaza is 3 star and as such I did not have great expectations, so was happy that it more than met our rei										
14	Dubai	Admiral Plaza Hotel	Horrible stay away!! Terrible staff terrible	Do not stay here at any cost. Having been to dubai a number of times before I wanted a hotel near the meetin										
15	Dubai	Admiral Plaza Hotel	Suitable for Stopover	I intentionally booked a 3 star, expecting an average room and service standard. After reading the reviews here										
16	Dubai	Admiral Plaza Hotel	Stay away...Far far away	Where do I start with this disaster of an establishment called the Admiral Plaza Hotel. To make things clear I did										
17	Dubai	Admiral Plaza Hotel	a wrongly advertised hotel	The hotel advertises itself as having a minibar (free but only stocked with water), an English pub but now discor										
18	Dubai	Admiral Plaza Hotel	This place is far really	I am finding it hard to reconcile some of the comments offered by others with my direct experience. OK, so ther										
19	Dubai	Admiral Plaza Hotel	better than an average hotel	this hotel is smaller than the other hotels I have stayed in but it was clean and the staff was accommodating.brea										
20	Dubai	Admiral Plaza Hotel	not worth the money	When we arrived we were told we're lucky to have such a nice room and all...what they described to be "nice"										
21	Dubai	Admiral Plaza Hotel	Hotel not as described	DO NOT STAY IN THIS HOTEL unless you are happy to do nothing but watch TV. Staying at this hotel ruined our										
22	Dubai	Admiral Plaza Hotel	Good (Plaza) Hotel	I booked my stay at the Admiral Plaza through Expedia. Goodroom Package. I chose it based on location and 3.4										

Figure 1 The preprocessed Dataset

After the above mentioned steps, the dataset is now ready for convenient information and feature extraction. These steps are carried out with the help of TextBlob, which is a python library used for processing textual data available in an unorganized form, in a dataset. It does so by providing a simple API (Application Programming Interface). Furthermore, the polarity of each review for every hotel across all cities is calculated. The sentiment analysis is also done by TextBlob. Subjectivity is also provided by TextBlob which refers to the meaning of the sentence or a review, whether the review is a personal opinion, emotional outburst or a public opinion or factual information. After the polarity is calculated, the sentiment is analyzed which basically is the determination of the attitude or emotion of the user while writing the reviews about the hotel, namely positive, negative or neutral.

Polarity Calculation And Sentiment Analysis: Polarity is a float value which usually lies in the range [-1,1], according to the polarity the sentiment is analyzed. In our work the range is calculated as follows:

Positive: If the polarity is above or equal to 1.1 then the sentiment is considered to be positive which is given by:
 if review['polarity_mod'] >= 1.1:
 review['sentiment'] = 'positive'

Negative: If the polarity is equal to or below 0.9, the sentiment will be analyzed as negative. It is given by:

elif review['polarity_mod'] <= 0.9:
 review['sentiment'] = 'negative'

Neutral: If the value is between 0.9 and 1.1, then the sentiment is assumed to be neutral. It is given by:
 else:
 review['sentiment'] = 'neutral'

Where polarity_mod is calculated by:

$$\text{review}[\text{'polarity_mod'}] = \text{review}[\text{'polarity'}] + 1.$$

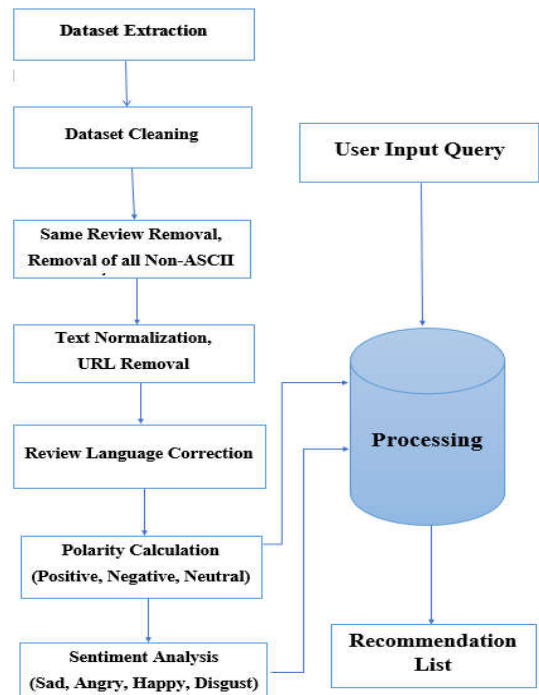


Figure 2 Proposed System Architecture

These steps when carried out yields out the recommendation list of hotels as queried by the user. This model has the potential to serve various business models which can implement this model to develop a business model and can also be used to expand the existing business models by providing valuable insights into the hotel industry. The proposed system is a well-constructed model which, the basic ingredients of the model is given in Figure-2 above.

RESULTS

As the proposed system is built using python, which is a high-level programming language considered to be very user-friendly language. As it is mentioned in the previous section the polarity calculation and sentiment analysis, polarity extracted is then converted to polarity_mod, which is finally used to generate ratings. Ratings are calculated in this manner:

$$\text{Ratings} = \text{review}[\text{'polarity_mod'}] * 2.5,$$

#	A	B	C	D	E
1	Dubai	ABC Almanar Hotel Apartment	0.157777778	2.894444444	positive
2	Dubai	Admiral Plaza Hotel	0.457003367	3.642508418	positive
3	Dubai	Admiral Plaza Hotel	0.266666667	3.166666667	positive
4	Dubai	Admiral Plaza Hotel	0.525	3.8125	positive
5	Dubai	Admiral Plaza Hotel	0.183585137	2.958962843	positive
6	Dubai	Admiral Plaza Hotel	0.375	3.4375	positive
7	Dubai	Admiral Plaza Hotel	0.142857143	2.857142857	positive
8	Dubai	Admiral Plaza Hotel	0.5	3.75	positive
9	Dubai	Admiral Plaza Hotel	0.46452381	3.661309524	positive
10	Dubai	Admiral Plaza Hotel	0.031944444	2.579861111	neutral
11	Dubai	Admiral Plaza Hotel	0.368627451	3.421568627	positive
12	Dubai	Admiral Plaza Hotel	-0.004724702	2.488188244	neutral
13	Dubai	Admiral Plaza Hotel	0.105448718	2.763621795	positive
14	Dubai	Admiral Plaza Hotel	-8.82E-05	2.49977944	neutral
15	Dubai	Admiral Plaza Hotel	-0.043181818	2.392045455	neutral
16	Dubai	Admiral Plaza Hotel	0.238938492	3.09734623	positive
17	Dubai	Admiral Plaza Hotel	0.390277778	3.475694444	positive
18	Dubai	Admiral Plaza Hotel	-0.131140351	2.172149123	negative

Figure 3 Generation of Results

In the above figure, the polarity is given in column 3, from which polarity_mod can be calculated from polarity finally resulting in rating which is computed by the formula given above.

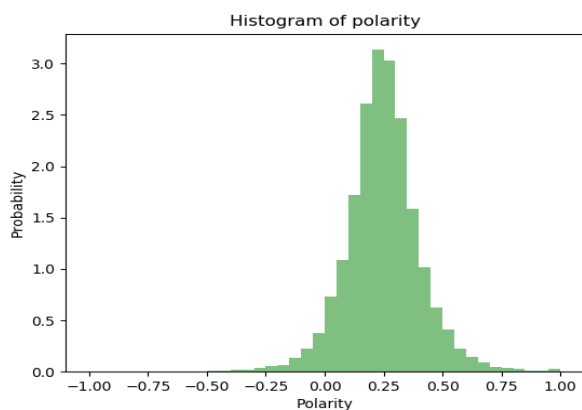


Figure 4 Polarity Histogram

The polarity for a particular city can be estimated by the steps mentioned in the previous section. In our case, the city taken into account is Dubai which contains around 240 hotels. We have created a histogram depicting the polarity calculation of a particular city. This plot is created using Matplotlib, which is a plotting library used for creating quality pictures in python. The plotting generated is illustrated in figure 4 given above. The sentiment analysis is also performed for this particular city, which is also portrayed in figure 5. It is visible enough that this system can be employed for various business models. It can influence heavily the business of any hotel which is looking to expand its reach. Sentiment analysis is shown in the figure below which clearly narrates the entire story i.e majority of reviews fall into the positive category which is an indication that the users have admired majority of hotels in this city.

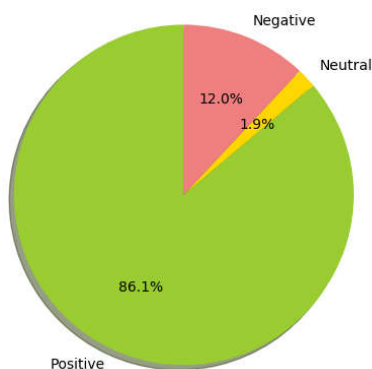


Figure 5 Sentiment Analysis

```
Python 2.7.14 (v2.7.14:84471935ed, Sep 16 2017, 20:19:30) [MSC v.1500 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:\Users\dell\Desktop\Final Thesis\Thesis\suggester.py =====
1. Shanghai
2. London
3. Dubai
4. New Delhi
5. Beijing
6. Chicago
7. Las Vegas
8. Montreal
9. New York
10. San Francisco

Input the city number: 3
Enter the least hotel rating you require in a range of 0 to 5: 2
Enter the maximum hotel rating you require in a range of 2 to 5: 4
Enter The Number Of Hotels: 10
Th Place Deluxe Hotel Apartments 3.3592680318017263
Beit Al Bahar 3.3637529366345236
The Baron Hotel Apartments 3.3691705684931352
Towers Rotana Dubai 3.3738727472787273
Savoy Park Apartments 3.3759747023809523
Grand Midwest Express Hotel Apartments 3.382084238099863
Starmetro Deira Hotel Apartments 3.385845958575738
Jumeirah Beach Club Resort And Spa 3.4089786414166374
Savoy Crest Exclusive Hotel Apartments 3.537954695787196
London Crown Apartments 2 3.875
>>>
```

Figure 6 Proposed Recommender System

The proposed recommender system is a remedy for researchers and travelers across the globe as it recommends hotels in major cities. The recommender system suggests hotels according to the user's preferences, enabling the user to select cities, and further selecting the range under which rating is selected. The user has the option to select rating range, on a scale of 0-5. The blueprint of the proposed recommender system is depicted in figure 6.

CONCLUSION AND FUTURE WORK

Recommender systems are currently being used in various domains whether be entertainment, tourism, books, education and many more. The scope of these systems is evolving at a rapid pace, mainly due to the information explosion that has taken place in recent times. The techniques behind these systems have improved significantly over the time. This study aims to provide a basic recommender system which renders suggestions to the users about the hotels located in major cities across the globe. The proposed system is fed with textual reviews according to which the recommender systems suggests top hotels based on the users input. The system calculates the polarity of the reviews based on which sentiment of the user can be analysed which can be used by start-up companies, existing companies can also benefit from this model as a tool to further extend their business. The ratings are calculated by the polarity, users have the privilege to select the range of desired ratings on a scale of 0-5, depending on this range and the city selected by the user the system recommends the most suitable hotels in the particular city.

This approach is a single criteria recommender system which purely recommends hotel based on the overall rating of the hotel. The emergence of multi-criteria recommender systems has the potential of providing accurate recommendations, as it focuses on multiple aspects of user's preferences resulting into more suitable results. This system can also consider multi-criterion recommender system in near future, taking more criterions such as food, cleanliness, connectivity with tourist places so as to capture more accurate preferences of the users. As this will result into additional data, this could be worked upon to produce better results. As the recommender systems are on rise, there could be systems which would recommend hotels by comparing prices of the same hotels on different websites, as there are contrasting prices of same hotels on various websites. There can be numerous other features to train a recommender system can be pondered upon. For example, user age can be a factor as recommender system will be able to recommend hotels along with club facility to the young users. It can recommend educational experts the hotels with library facility and so on. These all areas can be studied by the researchers to further develop this recommender system.

References

1. Chen, J., Chao, K., & Shah, N. (2013). Hybrid Recommendation System for Tourism. 2013 IEEE 10th International Conference on E-Business Engineering. doi:10.1109/icebe.2013.24
2. Dehkordi, Y. H., Thomo, A., & Ganti, S. (2014). Incorporating User Reviews as Implicit Feedback for Improving Recommender Systems. 2014 IEEE Fourth International Conference on Big Data and Cloud Computing. doi:10.1109/bdcloud.2014.51

3. Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*,16(3), 261-273. doi:10.1016/j.eij.2015.06.005
4. Kang, G., Liu, J., Tang, M., Liu, X., Cao, B., & Xu, Y. (2012). AWSR: Active Web Service Recommendation Based on Usage History. 2012 IEEE 19th International Conference on Web Services. doi:10.1109/icws.2012.86
5. Ketcham, M., Ganokratanaa, T., & Bansin, S. (2016). The Forensic Algorithm on Facebook Using Natural Language Processing. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). doi:10.1109/sitis.2016.103
6. Mcfee, B., Barrington, L., & Lanckriet, G. (2012). Learning Content Similarity for Music Recommendation. *IEEE Transactions on Audio, Speech, and Language Processing*,20(8), 2207-2218. doi:10.1109/tasl.2012.2199109
7. Meng, S., Dou, W., Zhang, X., & Chen, J. (2014). KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications. *IEEE Transactions on Parallel and Distributed Systems*,25(12), 3221-3231. doi:10.1109/tpds.2013.2297117
8. Sanchez, F., Alduan, M., Alvarez, F., Menendez, J. M., & Baez, O. (2012). Recommender System for Sport Videos Based on User Audiovisual Consumption. *IEEE Transactions on Multimedia*,14(6), 1546-1557. doi:10.1109/tmm.2012.2217121
9. Sharma, Y., Bhatt, J., & Magon, R. (2015). A Multi-criteria Review-Based Hotel Recommendation System. 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. doi:10.1109/cit/iucc/dasc/picom.2015.99
10. Shrote, K. R., & Deorankar, A. (2016). Review based service recommendation for big data. 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). doi:10.1109/aeicb.2016.7538334
11. Xing, Z., Parandehgheibi, M., Xiao, F., Kulkarni, N., & Pouliot, C. (2016). Content-based recommendation for podcast audio-items using natural language processing techniques. 2016 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2016.7840872
12. Yao, L., Sheng, Q. Z., Ngu, A. H., Yu, J., & Segev, A. (2015). Unified Collaborative and Content-Based Web Service Recommendation. *IEEE Transactions on Services Computing*, 8(3), 453-466. doi:10.1109/tsc.2014.2355842
13. Zhao, Z., & Shang, M. (2010). User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop. 2010 Third International Conference on Knowledge Discovery and Data Mining. doi:10.1109/wkdd.2010.54

How to cite this article:

Furkan Ahmad and Akheela Khanum M (2018) 'Review Based Recommender System For Business Advantage Using Natural Language Processing', *International Journal of Current Advanced Research*, 07(5), pp. 12445-12449. DOI: <http://dx.doi.org/10.24327/ijcar.2018.12449.2188>
