POI-BASED SERVICE RATING PREDICTION FOR PERSONALIZED RECOMMENDATIONS

Y. Pallavi[1], Dr. K. Madhavi[2]
PG Scholar[1], Assistant Professor[2]
CSE Dept[1,2], JNTUACEA[1,2], Anantapur, A.P, India[1,2]

Abstract: **Reviews on various services and products made available over Internet have been playing important role in helping people to make well informed decisions. Different services and their ratings are provided online so as to help prospective customers to understand the companies that offer better services. Therefore, online reviews became goldmine for researchers and academicians to build new algorithms and frameworks that can produce useful information to general public. Many researchers contributed towards mining online reviews to gain business intelligence. In this paper a framework has been projected that generates personalized Point of Interest (POI) based service recommendations using unsupervised learning methods. The proposed methodology has an underlying algorithm named POI Based Service Recommendations (PBSR) to achieve this. Datasets are collected from YELP portal. We built a prototype application to demonstrate proof of the concept. The empirical results revealed that the proposed system is capable of generating personalized service recommendations that are useful to prospective customers to make well informed decisions.**

*Keywords —**Service rating prediction; online reviews; recommendations; user point of interest.***

1. INTRODUCTION

Recommendations are there in every domain. It is more in online applications and social networks. Recommendations are also possible based on online reviews. As online reviews are gaining popularity in decision making, they are explored by the researchers to garner business intelligence.

Many recommender systems came into existence in numerous areas of businesses. For instance, in e-commerce, recommendations help users to make well informed decisions faster besides being aware of promotions and cost advantages. In the same fashion, service recommendations can help users or customers to make decisions while seeking hospitality, food, tourism and other services.

Recommendation systems studied in [2], [3], [4], [6] and [7]exhibit different methods used for recommendations. However, the one which is widely used is collaborative filtering. User preferences in providing recommendations play vital role in satisfying customers. The preferences of users are known as Points of Interest (POIs). These are used in this paper to make the recommendations on services.

We proposed an algorithm based on POIs to generate personalized recommendations. The formula makes use of unsupervised learning to group related reviews to support the user preferences. The proposed system predicts service ratings and generates personalized service recommendations. The recommendations are helpful to customers to have quick understanding on services and making decisions. The rest of the paper is structured as follows.

Section II provides review of literature. Section III presents the proposed system in detail. Section IV presents experimental results whereas section V concludes the paper and provides future scope.

2. RELATED WORK

This section provides review of literature involving mining of online reviews. Location based rating prediction model (LBRP) is evaluated in [1]. Location based social networks and the influence of geo-social on them is explored and evaluated in [2]. A methodology is proposed to have personalized recommendations based on social circle and user interests. Social similarities with adaptive approach are studied to generate recommendations in [3]. Item-level social influence prediction is investigated in [4] for users and posts rankings to know who should share and what to share in social networking. On-line social networks are studied in [5] to possess circle based recommendations. On the other hand social discourse recommendations are explored in [6]. While collaborative filtering technique is employed in [7] for achieving latent preference analysis in probabilistic way. Similarly collaborative filtering technique is employed in [8] for latent variable models based on Bayesian approach. In [9] also collaborative filtering algorithms are employed to have item-based recommendations. They provided user-based and item-based recommendations; Users' behaviors over online social networks are studied in [10] to create customized recommendations. Geo-social networking information is employed in [11] for generating preference-aware location based recommendations. A multi-faceted approach for collaborative filtering is explored in [12] while a matrix resolution method is providing recommendations in social networks and temporal dynamics besides collaborative filtering is studied in [13]. In [14] user photos are mined to have business intelligence on customized recommendations of products or services. Personalized ranking is employed to process interest growth for improved collaborative filtering in [15]. Completely different recommender systems are reviewed in [16]. Within the literature it's found that recommendation systems are used to facilitate users to create choices quicker. However, in this paper, we tend to target on generating personalized service recommendations supported by user preferences (points of interest) which create a lot of help.

3. PROPOSED FRAMEWORK

On-line reviews in social media and web sites like YELP are playing important role in influencing folks in decision making. During this context, we tend to plan a framework which will offer personalized service recommendations based on user-given points of interest.

Machine learning techniques concerning unsupervised learning is employed to achieve this. Clustering algorithm and other steps becomes the proposed algorithm to create personalized service recommendations. Online reviews, personalized service recommendations, online reviews YELP dataset provide rich data within the type of text.

Mining such text provides useful data that will be used to build choices with ease. The results of data mining offer needed business intelligence which will facilitate in creating well informed selections.

As shown in Figure 1, it's evident that the unsupervised learning that is clustering algorithm takes datasets from YELP and user-given points of interest. Then it performs clustering. Once clustering is completed, similar reviews are grouped together.

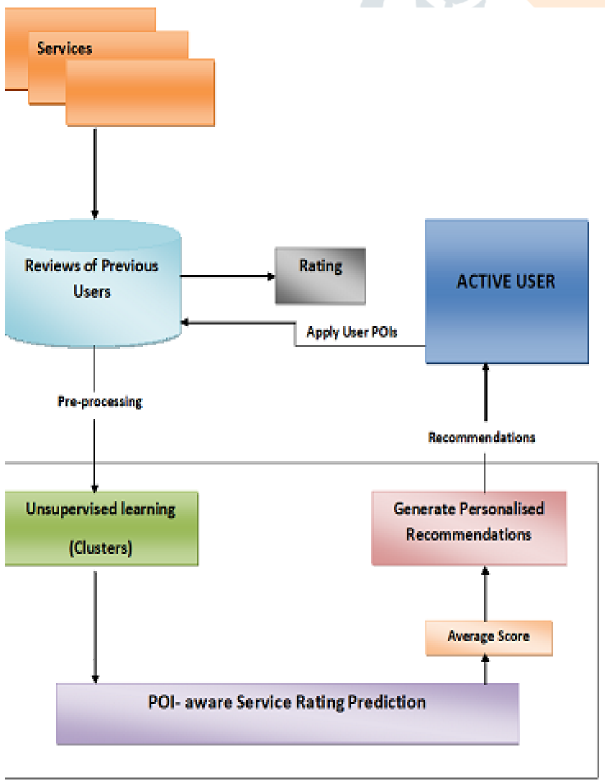

Figure 1: Framework for manufacturing personalized Service recommendations.

Based on the POI, the POI-aware service rating is expected and personalized recommendations are provided to end users. In the process two algorithms are proposed. The main rule thought as POI based service recommendation algorithm, which calls another algorithm known as clustering algorithm.

Algorithm: POI Based Service Recommendations (PBSR) Algorithm

---

Inputs: YELP Datasets, User POIs,
Output: POI based service recommendations
1. Initialize clusters vector C
2. C = Invoke Clustering()
3. For each cluster c in C
4. Predict POI-aware service rating
5. Generate personalized recommendations
6. End for
7. Return POI based service recommendations

Algorithm 1: POI base service recommendations algorithm

The POI based service recommendations algorithm takes online reviews from YELP and performs similarity based clustering. Cosine similarity is employed to find similarity between reviews. The results of similarity live are wont to build selections in grouping reviews. Once clustering is done, then the POI-aware service rating is expected so as to make personalized service recommendations.

3.1 Unsupervised Learning

Machine learning algorithms such as Clustering is employed to cluster reviews collected from YELP datasets. This algorithm takes care of clustering part. It is invoked by Algorithm 1. It takes variety of clusters and datasets as input and produces clusters of reviews.

Input: number of clusters k, Dataset D ,
Output: Clusters C
1. Randomly select k objects from D as initial cluster centres
2. Repea
3. Find similarity measure
4. Reassign object to cluster
5. Compute mean value of objects
6. Adjust cluster centroid
7. Until all objects are clustered

Algorithm 2: Clustering algorithm

Input: variety of clusters k, Dataset D.
Output: Clusters C at random choose k objects from D as initial cluster, centers repeat Find similarity measure, reassign object to cluster, compute mean value of objects, adjust cluster centroid Until all objects are clustered.

Algorithm 2: Clustering algorithm: The results of the formula are crucial to reduce area and time complexity besides rising performance of the projected framework. Once the clustering is completed, the remainder of the method is finished with less complexity. The method includes creating personalized service recommendations supported the ratings for given POIs.

3.2 Dataset Description

YELP dataset is considered as it is one of the famous shopping website, users across the world give ratings to services through the YELP.COM web site [21]. It has lots of users and their reviews on various services like restaurants, food and so on. The small point of datasets employed in this paper is as follows.

Dataset	Number of User	Number of Items	Number of Ratings
Food	9770	21370	341573
Restaurants	10449	67857	321551
Nightlife	11152	21647	436301
Shopping	8121	15460	112844

Table 1: Details of YELP datasets

As shown in Table 1, the datasets considered from YELP include food, restaurants, night life and shopping. It also shows variety of users, variety of things offered and also the variety of ratings related to every dataset.

4. IMPLEMENTATION DETAILS

A prototype application is made to demonstrate proof of the construct. The application is developed using Java platform. It provides intuitive interface for choosing datasets with online reviews, perform rating prediction, find ranking and at last create personalized recommendations.

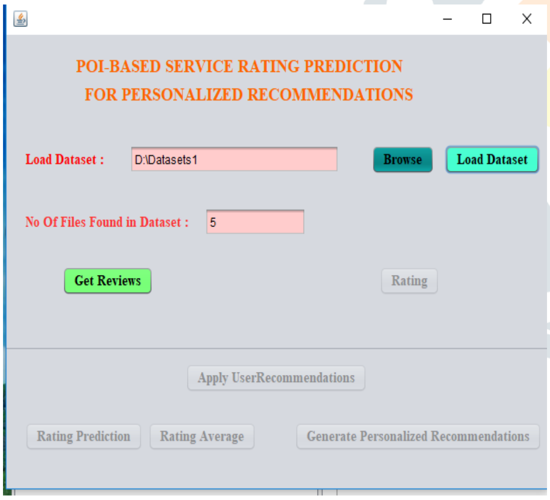

Figure 2: Shows main UI of the appliance

As shown in Figure 2, it's evident that the five datasets are loaded and the Get Reviews button is enabled that helps in presenting reviews that are part of the chosen datasets.

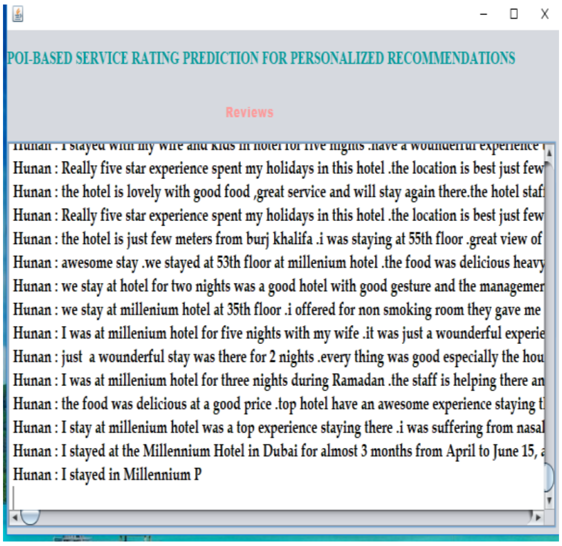

Figure 3: Shows reviews of selected datasets

As shown in Figure 3, the reviews found from the five chosen datasets are shown. These reviews are subjected to further analysis.

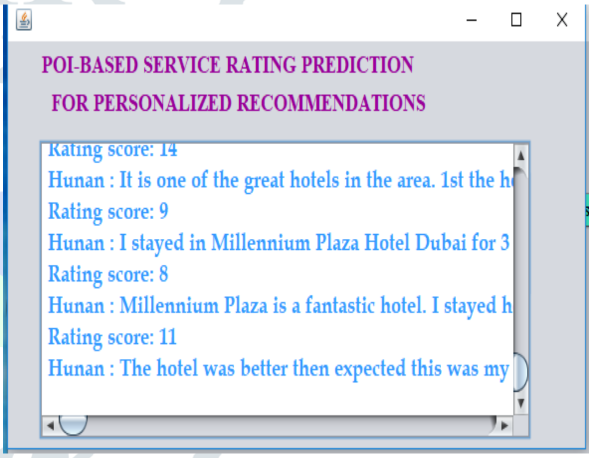

Figure 4: Shows computed rating

As shown in Figure 4, the rating computation is completed and also the results are shown. These are wont to generate average score for every service supplier.

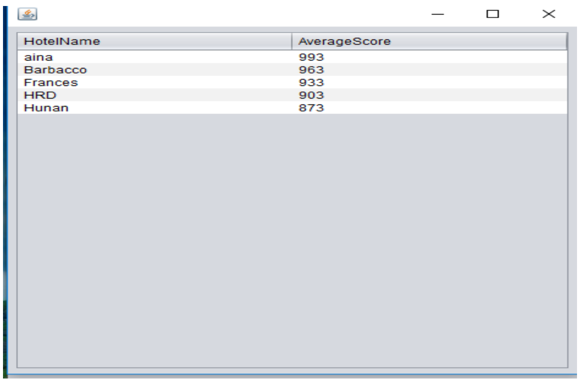

Figure 5: Average score of service suppliers (Hotels)

The typical score computed from the review ratings for every hotel is given in Figure 5. These details help in making personalized recommendations. It also helps new customers to make well informed decisions while preferring hotels to avail various services. The user preferences are considered by the appliance so as to possess additional correct recommendations.

5. EXPERIMENTAL RESULTS

Experiments are made to observe the potency of the projected methodology. The datasets collected from YELP internet site are used. A prototype application built using Java programming language with Graphical User Interface (GUI) to demonstrate proof of the concept. The experimental results discovered that the projected system is in a position to perform higher than the current system.

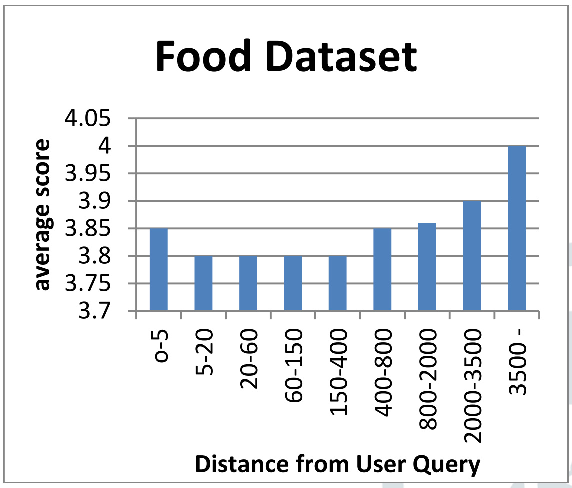

Figure 6: Average score vs. Distance from user question for food dataset.

As shown in Figure 6, it's evident that the horizontal axis shows the vary of distances from user question point. The vertical axis presents average score. Because the distance is raised, the typical score is raised step by step. Thus there is a relationship between the distance from user query and therefore the average score. These observations are made with food dataset.

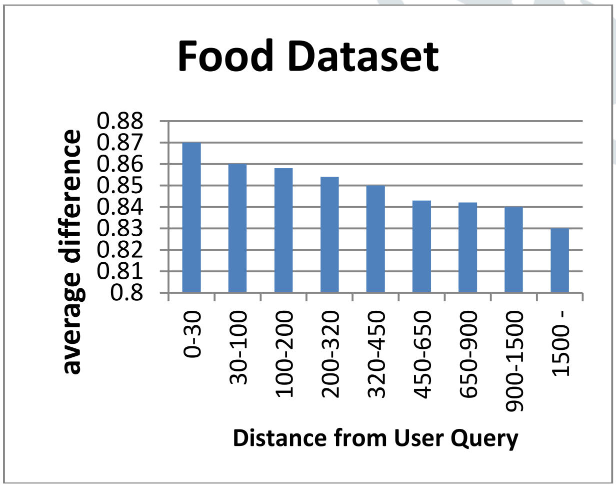

Figure 7: Average difference vs. distance from user query for food dataset.

As shown in Figure 7, it's evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average difference. As the distance is increased, the average difference is increased gradually. Therefore there is relationship between the distance from user query and the average difference. These observations are made with food.

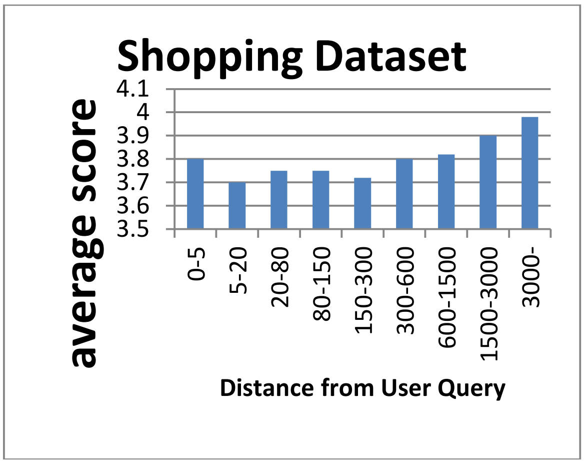

Figure 8: Average score vs. distance from user query For shopping dataset.

As shown in Figure 8, it's evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average score. As the distance is increased, the average score is increased gradually. Therefore there is relationship between the distance from user query and the average score. These observations are made with shopping dataset.

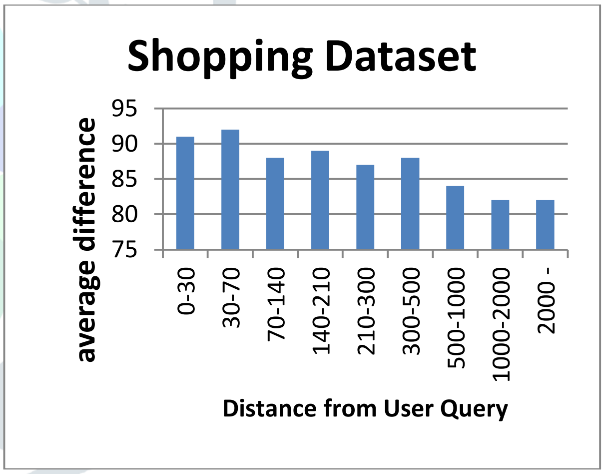

Figure 9: Average difference vs. distance from User query for shopping dataset.

As shown in Figure 9, it's evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average difference. As the distance is increased, the average difference is increased gradually. Therefore there is relationship between the distance from user query and the average difference. These observations are made with shopping dataset.

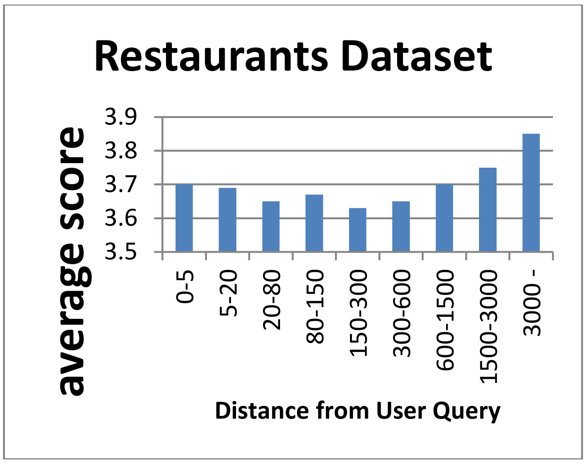

Figure 10: Average score vs. distance from user query For Restaurants dataset.

As shown in Figure 10, it is evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average score. As the distance is increased, the average score is increased gradually. Therefore there is relationship between the distance from user question and the average score. These observations are made with Restaurants dataset.

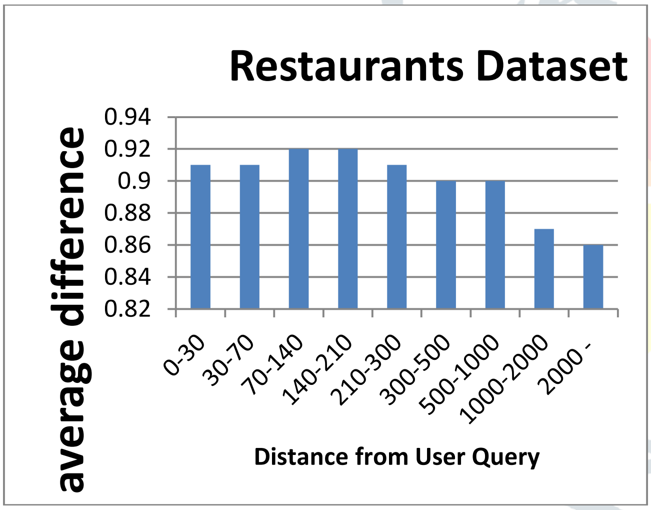

Figure 11: Average difference vs. distance from user query for Restaurants dataset.

As shown in Figure 11, it’s evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average difference. As the distance is increased, the average difference is increased gradually. Therefore there is relationship between the distance from user question and the average difference. These observations are made with Restaurants dataset.

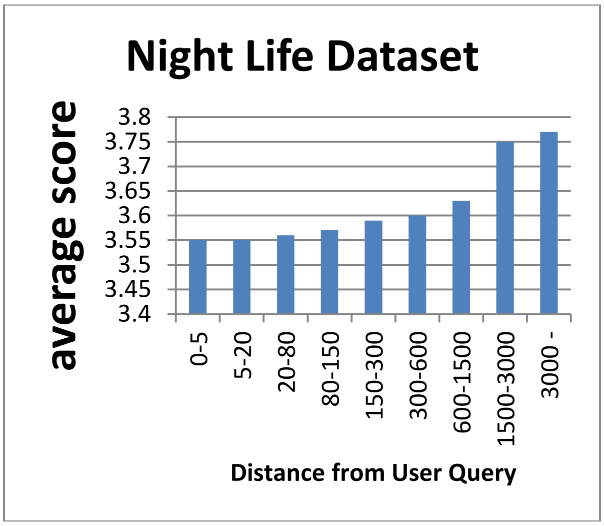

Figure 12: Average score vs. distance from user query For Night Life dataset.

As shown in Figure 12, it’s evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average score. As the distance is increased, the average score is increased gradually. Therefore there is relationship between the distance from user query and the average score. These observations are made with Night Life dataset.

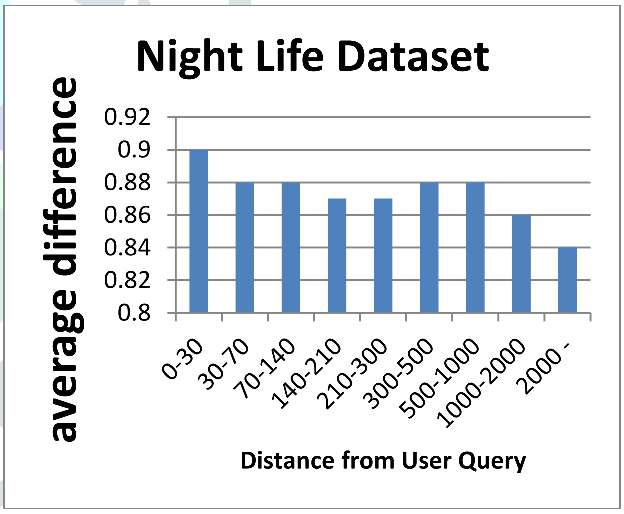

Figure 13: Average difference vs. distance from user query for Night Life dataset.

As shown in Figure 13, it’s evident that the horizontal axis shows the range of distances from user query point. The vertical axis shows average difference. As the distance is increased, the average difference increased gradually. Therefore there is relationship between the distance from user query and the average difference. These observations are made with Night Life dataset.

Yelp dataset	RMSE (Exist)	RMSE (Prop)	MAE (Exist)	MAE (Prop)
Food	0.996	0.886	0.769	0.689
Restaurants	1.053	1.005	0.851	0.796
Night life	1.126	1.05	0.897	0.8
Shopping	1.306	1.124	1.023	1

Table 2: Performance comparison with RMSE and MAE measures.

As shown in Table 2, the results of RMSE and MAE are shown for existing and proposed systems against four YELP datasets like food, restaurants, night life and shopping.

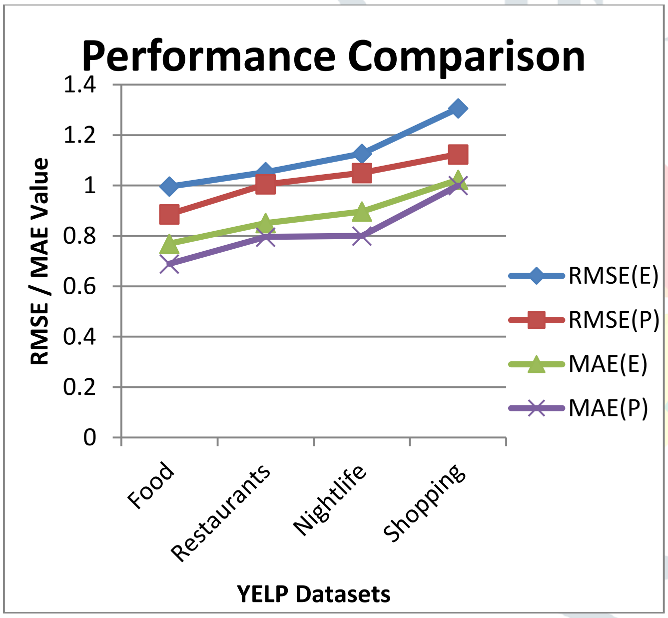

Figure 14: Performance comparison with existing system (RAME & MAE).

As given in Figure 14, it's evident that the RMSE and MAE are the employed to compare performance of the projected system with that of existing. The projected system showed less RMSE and MAE for all datasets. This is revealing the performance improvement over existing system. The YELP datasets provided in horizontal axis are compared with RMSE/MAE values given in the vertical axis.

6. CONCLUSION AND FUTURE WORK

In this paper, the proposed architecture to create personalized service recommendations using machine learning approach. Unsupervised learning such as clustering is used to group related reviews. Then service rating prediction is made to create personalized service recommendations. Datasets are collected from YELP for demonstrating the utility of the projected system. An algorithm is proposed to create personalized service recommendations. Two performance measures like RMSE and MAE are employed to calculate the performance of the projected system. The results discovered that the projected recommendation system is useful in providing service recommendations based on service rating prediction.

In future we intend to concentrate on the detection pretend reviews online to confirm that the fake online reviews don't influence decision of individuals.

REFERENCES

[1]Guoshuai Zhao, Xueming Qian, "Service Rating Prediction by Exploring Social Mobile User's Geographical Locations," IEEE Traction on Big data,2016.

[2] Xiaofeng Yang, Qian Yang, and Christo Wilson. Penny for your thoughts: Checking out the 50 cent party on sinaweibo. In ICWSM, 2015.

[3] Kyumin Lee, PrithiviTamilarasan, and James Caverlee. Crowdturfers, campaigns, and social media: Trailing and revealing crowdsourced manipulation of socialmedia. In in ICWSM, 2013.

[4] Ke Zeng, Xiao Wang, Qingpeng Zhang, Xinzhan Zhang, and Fei Yue Wang. Behavior modeling ofinternet water army in online forums. In Ifac World Congress, pages 9858–9863, 2014.

[5] Fangtao Li, Yang Gao, Shuchang Zhou, Xiance Si, and Decheng Dai. Deceptiveanswer prediction with user preference graph. In Meeting of the Association for Computational Linguistics, pages 1723–1732,2013.

[6] Becchetti, Luca, Castillo, Carlos, Donato, Debora, Leonardi, Stefano, Baeza-Yates,and Ricardo. Link-based characterization and detection of web spam. 2006.

[7] Benczr, Andrs, Csalogny, Kroly, Sarls, and Tams. Link-based similarity search to fight web spam. Airweb, 2006.

[8] cheng Chen, Kui Wu, S. Venkatesh, and Xudong Zhang. Battling the internetwater army: Detection of hidden paid posters. CoRR, abs/1111.4297, 2013.

[9] Magno G. Rodrigues T. Almeida V Benevenuto, F. Detective spammers on twitter. In unitedly, electronic communication, anti-abuse and spam conference (CEAS), 2:12, 2010.

[10] Thomas P. Lyons and John W. Maxwell. Astroturf: Social group lobbying and company strategy. Journal of Economic science & Management Strategy, 13(4):561–597,2004.

[11] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonalves, AlessandroFlammini, and Filippo Menczer. Detecting work and pursuit political abuse in social media. In International Conference on Weblogs and Social Media, Barcelona,Catalonia, Spain, July, 2011.

[12] Xia Hu, Jiliang Tang, and Huan Liu. Leverage data across media for trasmitter detection in microblogging. In SIGIR, 2014.

[13] Xia Hu, Jiliang Tang, and Huan Liu. Online social transmitter detection. In TwentyEighth AAAI Conference on Artificial Intelligence, 2014.

[14] Aifang Xu, Xiaonan Feng, and Ye Tian. Revealing, characterizing, and detecting crowdsourcing spammers: A case study in community q and a. In INFOCOM, 2015.

[15] James Hoggan and Richard Littlemore. "climate cover-up: The crusade to deny international warming". Energy & Environment, 21(3):363–364, 2010.

[16] Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. Social transmitter detection with sentiment information. In IEEE International Conference on Data Processing, pages 180–189, 2014.

[17] Yelp (2018). Retrieved from https://www.yelp.com/. Accessed on 20-3-2018.