Hanapresto: A Platform for Restaurant Businesses with Recommender System using Knowledge Extraction from Social Media and Customer Preferences

Asia Pacific Journal of Multidisciplinary Research
Vol. 7 No.2, 97-105
May 2019
P-ISSN 2350-7756
E-ISSN 2350-8442
www.apjmr.com
CHED Recognized Journal
ASEAN Citation Index

Henry Dyke A. Balmeo[1], Albert A. Vinluan[2]
[1]Information Technology Department/Graduate School, University of the East, Manila, Philippines, [2]College of Computer Studies, New Era University, Quezon City, Philippines
henrydyke@yahoo.com[1], aavinluan@neu.edu.ph[2]

Date Received: April 5, 2018; Date Revised: March 5, 2019

Abstract – *The restaurant and mobile food performance had the highest number of company which offers nourishment benefit due to extraordinary number of consumers. With the massive number of restaurants in Manila and incredible number of customers who liked to eat out, there is issue of finding the best restaurants to eat. In association with that clients set aside plenty of opportunity to search for the restaurants that suite their preferred budget. The study expects to develop an application for small and medium enterprises (SME's) that locates restaurants through Global Positioning System, and ranks restaurants and posts the best performing in view of social media reactions. The administrator of the system produced measurable report and positioning from extricated information from the web-based social media. The system used only using web and mobile platform. The framework essentially helped the clients and the restaurants since the application has capacities intended to satisfy both entities.*

Keywords – *Restaurant, Global Positioning System (GPS), Manila, Recommender System, Knowledge Extraction*

INTRODUCTION

Mobile technology or mobile devices have created impacts to the lives of many. This technology brought different services that help many people to ease the way they are living. The powerful smartphone and other mobile devices have given birth to lots of social media applications and many more in the network. [1]

The availability of internet and its web services like social media Facebook, Twitter, Instagram brought life to the business industry. It created channels for business to connect with their customers. The high level of use and interaction of Social Media influences greatly the business environment which is thus exposed to a paradigm shift, where hierarchies fall apart and the communication and collaboration create wider and wider networks for the employees and all the partners of the organizations [2] and it is supports the fact that social media became an essential part to prolong the performance of a business [3]

The presence of social media also helps Small and Medium Enterprise (SME) to compete with other big businesses. Social media help SME to easily advertise and market their products. Feedbacks from social media became valuable resources to innovate and improve their business processes [4].

Among SME, restaurants are on the top. At present, Metro Manila is on the top when it comes to the number of restaurants or establishments that provide food services. Accommodation and Food Service Activities for Establishments conducted nationwide preliminary results showed that there were a total of 5,475 establishments with total employment (TE) of twenty (20) and over engaged in accommodation and food service activities in the formal sector of the economy. Restaurants and mobile food service activities had the highest number of establishments at 3,956 or 72.3 percent, followed a far by short term accommodation activities with 1,113 establishments or 20.3 percent [5].

In the Philippines, the proliferation of vast arrays of food service facilities such as conventional full-service and fast food restaurants, coffee shops, food courts, roadside stalls, canteens, delicatessens, etc., together with improved purchasing power, growing time constraints among household members and incessant bombardment of promotional ads across

various media collectively create a strong impetus among Filipinos to "eat out" [6].

Report reveals a 13% decrease in monthly grocery spend of Filipino reported in 2014 compared to 2012. From a monthly spend of P5,400 in 2012, Filipino customers only spent P4,700 on an average in 2014. Driving this cut in grocery spend is the spike in the number of regular clients who are taking out and the increased regularly in eating out of home. In the information, 25% of consumers eat outside in any fast food restaurants at least once a week, and a contrast to a year ago with only 14% [7].

With the great number of restaurants in Manila and great number of customers who preferred to eat out, there is problem of finding the best restaurants to eat. In connection to that, customers take lots of time to look for the restaurants that suit their budget.[8] Since mobile devices and social media created new ways to discover new things, developing an application where social media can be used by the consumer to utilized it function to find restaurants to eat out [9].

The study aims to develop a mobile and web-based application to help the SME's owners promote their business and customers' needs in finding restaurants. The application for this web and mobile is different from the other existing applications because it is developed for the SME that has no website and with a small capital for a new in the business [10]

CONCEPTUAL FRAMEWORK

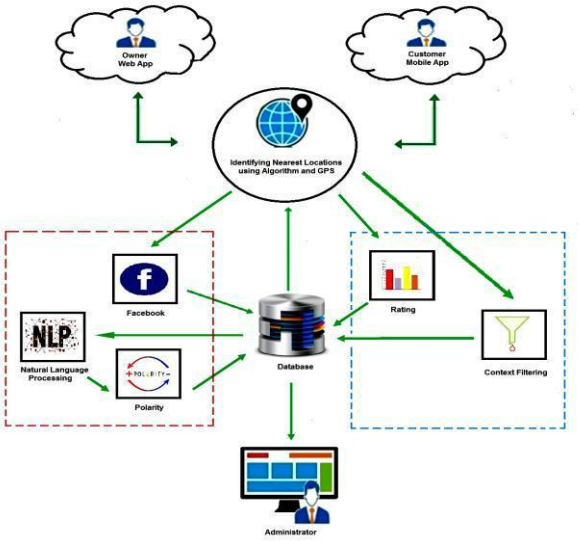

Fig. 1. Proposed Conceptual Framework

The innovation of Internet-based technologies such as mobile and web application by SME is important to evaluate the personality of the owner-manager given the moderating effect of managers on the innovations and technology adoption of the business. With these, the proposed system incorporates the mobile and web application to social networking site in promoting the SME's businesses as presented in Figure 1. The administrator has the capabilities to approve the owners request to promote their business, generate reports, update the system, and manage the information in the mobile and web application. Users/clients will register to the system to access, view, and rate and give feedback to different registered restaurants. The feedbacks/comments from the Facebook given by the clients will be processed by social media analysis tools to produce statistical reports.

OBJECTIVES OF THE STUDY

There are two approaches used in this study; Nearby Algorithm and Social Media Extraction. This system aims to develop an application that would suggest and help to find the users a restaurant by using Google maps API where the service is good and the actual product by displaying the rates for each restaurant according to users that previously experienced the service. The owner of the restaurants needs to register online through web to validate and accept the request by the administrator. The information about of the restaurants is posted in the Facebook's for purposes of advertising. The users comment and rate the restaurants based from data on Facebook page.

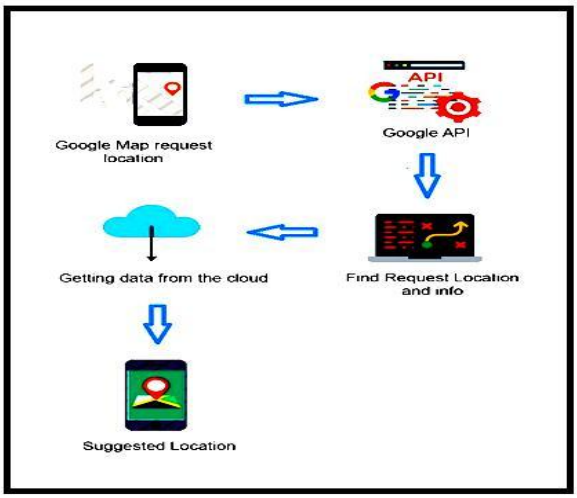

Fig. 2. Flow Extraction - Google maps

Figure 2 shows Google map API, the request to search for restaurants location or suggest the restaurants nearest location. Location request is sent using mobile application and tries to communicate with the Google maps, after validating the location, the mobile application identifies the nearest location and information of the restaurant and displays the information of the restaurant. The cloud or the internet storage is main depository of the database of the system.

MATERIALS AND METHODS

Hanapresto: A Platform for Restaurant Businesses with Recommender System using Knowledge Extraction from Social Media and Customer Preferences is intended for the use of the restaurants in Manila and its customers. The study used developmental research is different from the design-based research. This research emphasizes the study of learning as a result of designing unique instructional interventions. [11] Developmental studies often are structured in phases. It may have an analysis phase, design phase, a development phase, and a try-out and evaluation phase. Another would include phases directed toward first analysis, then prototype development and testing, and finally prototype revision and retesting.

Research Design And Framework

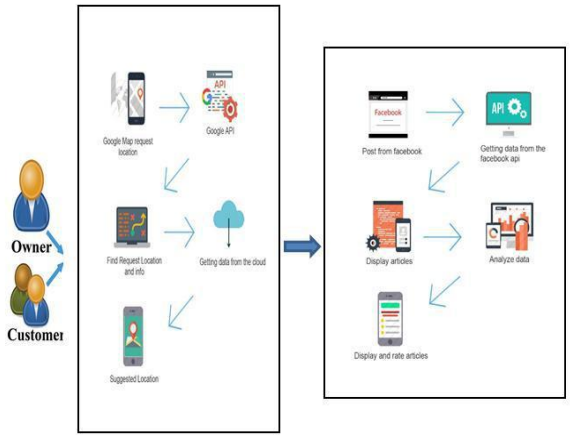

Fig. 3. Overview of the Project Methodology

This study as presented in Figure 3, is intended for the use of restaurants in Manila and its customer. Developmental studies, like this study often are structured in phases. It may have an analysis phase, design phase, a development phase, and a try-out and evaluation phase. Another would include phases directed toward first analysis, then prototype development and testing, and finally prototype revision and retesting.

The framework shows that there are three entities/users that use the system the customers and the restaurant owners. These users access the application through internet connect. A mobile and personal computer are the two platforms that users utilize. On the user side Global Positioning System is the main capability that they use. This capability helps them to locate restaurants that they preferred. Meanwhile, owners get data based on social media extractions. Reports through graphs were extracted from social media. Social Media extractions happen through social media crawler (directly connected to social media) then, through API request preprocessing of data occurs. Afterwards, it is stored to the database of the system for data extraction. Extracted social media data is stored to the main database. [12]

Software Development Process

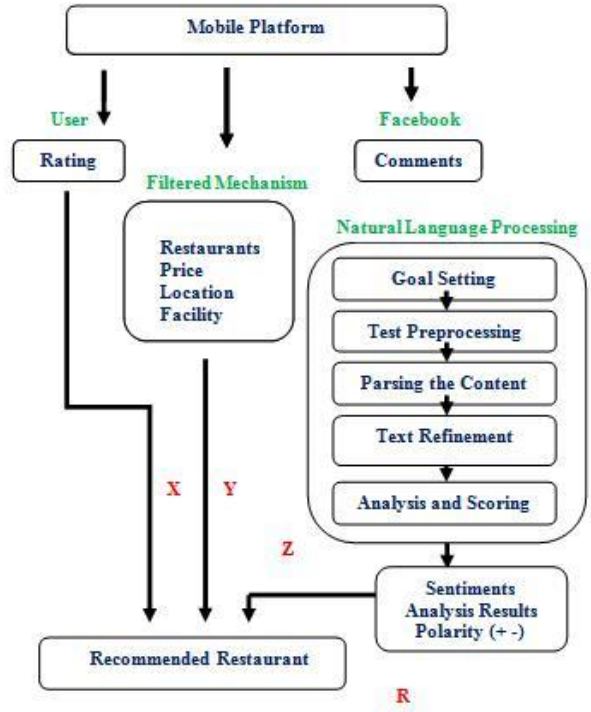

Fig. 4. Hanapresto Architecture

In Figure 4 showing Hanapresto Architecture, the diagrams identify the following process. To develop a mobile platform that identifies the rating base on the survey of the users and customers of the restaurants, the next process is to process the filtering mechanism with filtered the rating and results based on the restaurants, the locations, different food prices and restaurant facilities. The individual Facebook user account as well as and customers use the application

to comment, rate and survey the selected restaurant. The NLP extracted all the data and information especially the comments to compute and identify the positive and negative polarity in determining the sentiments analysis results. The following step in the sentiment analysis is explained in the flowchart as shown in Figure 4.

1. **Goal Setting.** It means to determine the sentiment analysis goal and the scope for the text content.
2. **Text Processing.** This involves determining the source i.e. whether you are taking the data from web, micro-blogging site, etc. The text is loaded to the processing system (the system, technique to be used for the analysis), unnecessary words from the text are deleted and organizing the emotional symbols that people use in texts into words and organized. Also, it observed that to express strong sentiments, uppercase alphabets are used (example: EXCELLENT).
3. **Parsing the Content.** It involves segmenting the words based on their polarity; tagging the parts of speech used (adjective, noun, etc.); identifying the terms.
4. **Text Refinement.** This is to guarantee the correct analysis and text refinement. It should be finds the stop words and synonyms, and others.
5. **Analysis and Scoring.** It involves determining the sentiment bearing phrases from the data and scoring the data result. Scoring is the process in which the intensity of the sentiment is analyzed.

RESULTS AND DISCUSSION

Table 1. Computed Longitude and Latitude with Geohash

Restaurant Name	Longitude	Latitude	Geohash
Sample 101	121.38210296631	14.260722942045	Urfpuxbrfzvx
Moon Leaf HB	120.9979683	14.6146617	Urfruxfpfrcx
Kevin's Bar	121.38330459595	14.26371761098	urfpuxfxbxyx
Hanapresto	120.981	14.5894	urfrgzzxvxvp
Hunger Buster HB	120.99469	14.56346	urfrgruzcpyz
E-GOV	121.9678	14.5865	urfrgzvxfpfx
Kevin's Restaurant	121.38278961182	14.261055685002	urfpuxbxfxvz
COUCH HB	120.981	14.5894	urfrgzzxvxvp
Gumbo HB	122.762	14.9846	urfxuzzryrzx
Chun Chon HB	120.98735	14.60783	urfrurvpgpgx
El Franco Resto	121.47720336914	14.3675081481	urfpyzypgpyx
Pat's Café	121.48097991943	14.363350796927	Urfpyzgpfrfx

Table 1 shows that the computation is based on the Haversine formula which is indicated in the algorithm, the output is in kilometers, and then converted into meters to produce the precise output in the mobile.

The researcher attempted to enhance the algorithm by determining which process restricted or limited the search for nearby-neighbors of a certain location. This chapter involves methods on how the researcher resolved the problems of the study by explaining further by the following observations done in a controlled environment:

In the Enhanced Algorithm, the researcher used the centralized locating method that created a new coverage area in locating the nearby establishments using the input coordinates as the center of the bounding box. The enhanced algorithm created a rectangle called the bounding box, based on the input of the user and took the coordinates of the vertices of the rectangle. Then, the Northwest and Southeast of the vertices are collected to produce another character for a new Geohash code that determines locations bounded by the boundary box.

Using the inputted data, latitude, longitude, and distance, the algorithm formulated a bounding box that located all nearby establishments of the inputted coordinate. Using the following formulas, the new bounding box was created:

*maxlat = lat + ((distance/6371) * (180/p));*
*minlat = lat - ((distance/6371) * (180/p));*

maxlng = lng + ((distance/6371/
cos(lat(p/180)) * (180/p));*

minlng = lng – ((distance/6371/
cos(lat(p/180)) * (180/p));*

These values serve as the foundation in creating the bounding box. By alternating these values, one pinpoints the corners of the bounding box.

North-East Point = NE(maxlat, maxlng)
North-West Point = NW(minlat, maxlng)
South-East Point = SE(maxlat, minlng)
South-West Point = SW(minlat, minlng)

After the bounding box has been created, if the bounding box crossed the boundary line, the algorithm creates a new shared prefix for the geohash of the establishments. The algorithm gets the geohash of the

longitude and latitude of the four corner points. The first values of the corner points constructed the new prefix. Any establishments inside the bounding box were considered as s nearby-neighbor of the inputted coordinate. These are the rules to consider if the establishment is inside the bounding box:

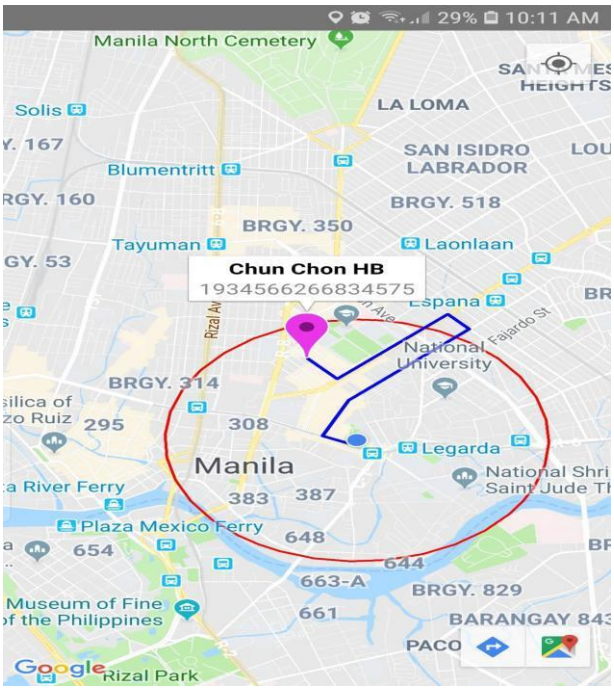

Fig. 5. Geohash Boundary Box

Figure 5 shows that the output of the existing algorithm's corresponding altitude suffix differentiates establishments on the same location but different in altitude. Furthermore, none of the Enhanced algorithm's first output was modified or changed when combining the suffix with the geohash code.

By adding the centralized locating method to include a new bounding box, nearby establishment that are previously cannot be located because of the boundaries can now be specified as a nearby establishment since the scope of the area is now determined by the user. Locations that are previously separated by the boundaries will now share a prefix that will determine that they are near each other. To further specify the establishments and to avoid overloading of data, the researchers added the function to choose what type of establishment will be searched for.

Table 2 shows the Hanapresto customers review that are stored and extracted from the databases. The extracted data composed of the id facebook account, comments of the customers, customer name, restaurant name, and the rating. The data were extracted from the mobile and web application where the users and customers placed their comments and ratings on the different categories of the restaurant.

Table 2. Hanapresto Customers Review based on Facebook page

ID	Comment	Customer_Name	Restaurant_Name	Rate
1	pwede mag vape ehhh	HENRY DB	SAMPLE 101	4
2	food	MARLON CARPENA	PAT'S CAFÉ	5
3	nice food	MARLON CARPENA	KEVIN'S RESTAURANT	5
4	Okey	MARLON CARPENA	KEVIN'S BAR	5
5	sana okey	MARLON CARPENA	PAT'S CAFÉ	5
6	good food	MARLON CARPENA	GUMBO HB	5
7	Good	MARLON CARPENA	GUMBO HB	5
8	Nice	MARLON CARPENA	SAMPLE 101	5
9	good menu	MARLON CARPENA	EL FRANCO RESTO	5
10	very nice	MARLON CARPENA	KEVIN'S BAR	5
11	Awesome	MARLON CARPENA	HUNGER BURGER HB	5
12	need to improve	MARLON CARPENA	HUNGER BURGER HB	5
13	Good	MARLON CARPENA	HUNGER BURGER HB	5
14	'very nice place	MARLON CARPENA	HUNGER BURGER HB	5
15	dapat may iba pang menu	MARLON CARPENA	HUNGER BURGER HB	5
16	happy to visit this resto	MARLON CARPENA	HUNGER BURGER HB	5
17	ang sherep!	HENRY DB	HUNGER BURGER HB	5
18	happy place	MARLON CARPENA	GUMBO HB	5
19	the best eto	MARLON CARPENA	HUNGER BURGER HB	5

Table 3 are the Hanapresto customers rating which are extracted from the stored database. This table shows the customer's id, rating, category of foods, facility, customer's name, and restaurant name. The mobile and web application extracted the data from the customers who rated the foods, facility, location and the restaurant.

Table 3. Hanapresto Customers Rating based on Customer review

ID	Rate	Category	Customer_Name	Restaurant_Name
1	3	FOOD	JOHN ROSH BIRADOR	SAMPLE 101
2	3	Food	HENRY DB	SAMPLE 101
3	1	Food	HENRY DB	SAMPLE 101
4	3	Facility	HENRY DB	SAMPLE 101
5	5	Food	MARLON CARPENA	HUNGER BURGER HB
6	3	Facility	MARLON CARPENA	KEVIN'S BAR
7	3	Food	MARLON CARPENA	PAT'S CAFÉ
8	1	Food	MARLON CARPENA	GUMBO HB
9	3	Facility	MARLON CARPENA	GUMBO HB
10	5	Food	MARLON CARPENA	SAMPLE 101
11	1	Facility	MARLON CARPENA	EL FRANCO RESTO
12	3	Food	MARLON CARPENA	KEVIN'S BAR
13	5	Food	MARLON CARPENA	HUNGER BURGER HB
14	5	Facility	MARLON CARPENA	HUNGER BURGER HB

Sentiment Analysis

Identifying the Individual Sentiment Score. The goal of the researcher is to improve text based sentiment identification from a single user by incorporating user similarity knowledge within a social network. To achieve this, the researcher first computed the individual text based sentiment scores of all entities by each user in the network. For individual scoring, the researcher used the textual sentiment algorithm. Three possible outcomes of the textual sentiment algorithm are "-1" (i.e. negative sentiment), "+1" (i.e. positive sentiment) and, "0" (i.e. neutral sentiment). The outcome of this step is a "user-entity" matrix $A \in \mathbb{R}^{nxm}$, where n is the number of users in the network and m is the number of entities observed. The element a_{ij} in Arepresents the sentiment score of the i^{th} user (u_j) on the j^{th} entity (e_j).

Table 4 shows the scoring per comments. The tabulated results are processed based on the following discussions. The comments gathered from the database folders and converted to the .csv file. The sentence splits into single words. The researcher compared the individual words of the sentence with database of words. In comparison, one finds the probability of both positive and negative labels. According to the comparison find the polarity. It is noted that positive comments scored dominated. The compound score is aggregated value for the three other scores, which defines the overall score of all the particular comment. The three other scoring was used for each token.

Table 4. Sentiments Analysis Table

ID	Comments	Stopwords	Negative	Neutral	Positive	Compound	Analysis
1	nice food		0	0.263	0.737	0.4215	-100 very negative/ serious
2	good food		0	0.256	0.744	0.4404	-100 very negative/ serious
3	Good		0	0	0	0.4404	100 very positive / enthusiastic
4	Nice		0	0	0	0.4215	100 very positive / enthusiastic
5	good menu		0	0.256	0.744	0.4404	100 very positive / enthusiastic
6	very nice		0	0.244	0.756	0.4754	100 very positive / enthusiastic
7	Awesome		0	0	1	0.6249	-14.2 Somewhat negative/ serious
8	need to improve	to	0	0.408	0.592	0.4404	-100 very Negative /serious
9	Good		0	0	1	0.4404	100 very positive/ enthusiastic
10	very nice place		0	0.393	0.607	0.4754	100 very positive/enthusiastic
11	happy to visit this Resto	this, to	0	0.519	0.481	0.5719	-100 very negative/ serious
12	happy place		0	0.213	0.787	0.5719	100 very positive/ enthusiastic
13	the best eto	the, eto	0	0.323	0.677	0.6369	100 very positive/ enthusiastic

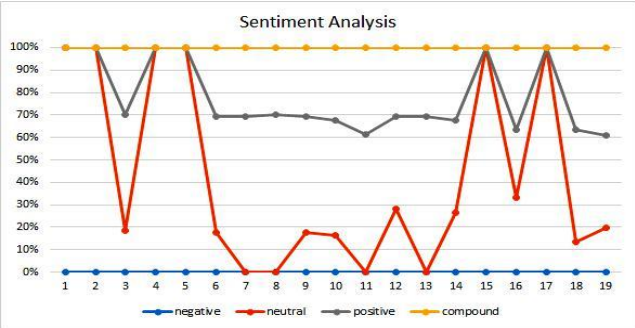

Fig. 6. Sentiment Analysis Graphical Presentation

Figure 6 presents the graphical presentation the sentiment scoring. It shows that compound score of every comments dominates the three other scoring, negative, neutral and positive polarity. The sentence percentage was rated according to their sentiment polarity towards their subject. That way, each comments was assigned to a positive, neutral or negative category, helping the researcher the machine learning algorithms used. After this categorization, there were almost 80% positive comments, 15% neutral comments and 5% negative comments. This minor tendency towards negativity affected some of the algorithms, either causing increased or decreased accuracy. In order to increase the effectiveness of the categorization on the experimentation, the comments had to be pre-processed. The pre-processing consisted of removing special characters that added no value to the sentiment polarity, such as the '#' character. The whole text of every tweet was converted to lower case characters and every web address in it was replaced by the keyword URL, since the actual link was of no importance, the important fact was that there was a link. As a last step the references to other users, using the '@' character, were replaced by the REF keyword since the username of the referred user had no impact on the sentiment polarity of the comments.

Computation:

User rating. The computation is based on the actual data gathered, summing up and dividing the total number of users who rated on the different category as shown in equation 1.

Formula:

$$X = (a \ / b) \qquad (1)$$

where

X = User Rating
a = Rating value

b = Total no. of users who rated

Filtered mechanism. The results are computed based on the following category:; restaurants, price, location, facility, as shown in equation (2) and (3).

Formula:

$$Y = (rr + rp + rl + rf) / n \qquad (2)$$

Then

$$Y = (rr + rp + rl + rf) / 4 \qquad (3)$$

Where

Y = Filtered mechanism
rr = restaurant rating
rp =price rating
rl = location rating
n = total no. of category

Restaurant Rating = Rating value / Total no. of users who rated

$$rr = (rv / rb) \qquad (4)$$

where

rr = restaurant
rating rv =rating
value

rb = Total no. of users who rated

Price Rating = Rating value / Total no. of users who rated

$$rp = (rv / rb) \qquad (5)$$

where

rp = price rating
rv =rating value
rb = Total no. of users who rated

Location Rating = Rating value / Total no. of users who rated

$$rl = (rv / rb) \qquad (6)$$

where

rl = location rating
rv =rating value
rb = Total no. of users who rated

Facility Rating = Rating value / Total no. of users who rated

$$rf = (rv / rb) \quad (7)$$

where

rf = location rating

rv =rating value

rb = Total no. of users who rated

Sentiment Analysis Result. Computation is based on the actual result in Table 4.3, Sentiment Analysis table.

Recommended Restaurant

Formula:

$$R = (X + Y + Z) /3 \quad (8)$$

where

X = User Rating

Y = Rating value

Z = Sentiment Analysis Result

After gathering and computing the following formula, the highest average result is considered the first and top recommended restaurant.

CONCLUSION

The spatio-temporal algorithm identifies the nearest location of the restaurant. The system provides location of the different restaurants based on the location through the global positioning system which basically helps the customers and restaurants since it is designed to satisfy both entities.

The Hanapresto application extracted data and information based on the customers review and rating. The extracted databases are composed of the facebook id account, comments of the customers, customer's name, restaurant's name and the rating. They are all extracted from the mobile and web application. The tabulated result on the system is based on the algorithm and the sentiment analysis which are provided on the system.

The recommendation was based on the different parameters provided by the users and the researcher. A series of analysis and computations was provided to identify the functionality of the recommender system.

RECOMMENDATION

Inasmuch as the current study focuses only on restaurant locations, physical facilities and amenities, price quotations of offered menus, the researcher recommends that similar studies be made to include a more wide-ranging features, such as fine dining, cuisine, delivery purchase and use of credit cards, exotic food and cuisine to satisfy every customer's preference and satisfaction.

Inasmuch as the current study is made operational only within Metro Manila, the researcher recommends that the system application produces a wider geographic coverage to include selected cities in NCR.

Inasmuch as the current study uses mobile devices, specifically, Android, the researcher recommends that future studies could utilize not only Android devices but also the Apple iOS and other new smart phone operating system.

The researcher recommends that a triplication of the current study be made to validate its findings.

REFERENCES

[1] Yadav, M., Joshi, Y., & Rahman, Z. (2015). Mobile social media: The new hybrid element of digital marketing communications. *Procedia-Social and Behavioral Sciences*, *189*, 335-343.

[2] Georgescu, M., & Popescul, D. (2015). Social Media–the new paradigm of collaboration and communication for business environment. *Procedia Economics and Finance*, *20*, 277-282.

[3] Paniagua, J., & Sapena, J. (2014). Business performance and social media: Love or hate? *Business horizons*, *57*(6), 719-728.

[4] Anderson, E. W. & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science 12*(2), 125-143.

[5] 2013 Annual Survey of Philippine Business and Industry (ASPBI) - Accommodation and Food Service Activities for Establishments with Total Employment of 20 and Over. Retrieved from https://www.psa.gov.ph/content/2013-annual-survey-philippine-business-and-industry-aspbi-accomodation-and-food-service

[6] Rufino, C. C. (2015). The Pattern of Consumption for Food Away From Home (FAFH) of Filipino Households duringthe Modern Era.

[7] Nielsen Shopper Trends Report, Fiipino Shoppers want more, more more retails channels.http://www.nielsen.com/ph/en/insights /news/2016/Filipino-shoppers-want-more-more -more-retail-channels.html

[8] Weinberg, B., and Pehlivan, E. (2011). Social spending: Managing the social media mix, Business Horizons, 54, 275-282.

[9] Rust, R. T. & Zahorik, A. J. (1993). Customer satisfaction, customer retention, and market share. *Journal of retailing*, *69*(2), 193-215.

[10] Terziovski, M. (2010). Innovation practice and its performance implications in small and medium enterprises (SMEs) in the manufacturing sector: a resource-based view. *Strategic Management Journal*, *31*(8), 892-902.

[11] Rufino, C. C. (2015). The Pattern of Consumption for Food Away From Home (FAFH) of Filipino Households during the Modern Era.

[12] Richey, R.C., Klein, J., & Nelson, W. (2004).Developmental research: Studies of instructional design and development. In D. Jonassen (Ed.) Handbook of *Research for Educational Communications and Technology (2nd ed.) (pp. 1099- 1130). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.*