Collaborative Filtering and Regression Techniques based location Travel Recommender System based on social media reviews data due to the COVID-19 Pandemic

K.V.Daya Sagar¹, Smt. P.S.G.Arunasri², Smt.Sridevi Sakamuri³, Smt.J.Kavitha⁴, Dr.DBK Kamesh⁵
¹Assoc. Professor, Dept. of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India - 522502.
²Assoc. Professor, Dept. of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India - 522502.
³Asst. Professor, Dept. of Electronics and Computer Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India -522502.
⁴Asst. Professor, Malla Reddy Engineering College, Hyderabad, Telangana,India.
⁵Professor, Dept. of Computer Science and Engineering, Malla Reddy Engineering College, Hyderabad, Telangana,India.

E-mail:sagar.tadepalli@gmail.com

Abstract
Nowadays, Dynamic industries like tourism is a enhance to boost many countries economy in recent years. The hotel sector leads to a significant role among all aspects of the tourism industry. Online travel platforms, in association with hotel management, are a part of hotel E-tourism that helps users to make travel plans online, suggest precise recommendations in consideration with the earlier feedbacks upon hotel stay. From the past few months, the user roaming ratio falls rapidly due to the COVID-19 Pandemic across the globe. In a business concern, this Pandemic affects all tourism sectors, specifically in the hotels’ occupancy. To retain business positions and attract the users in these problematic scenarios, hotel and travel platform management concerns about past preferences of users, which have a significant impact on gratification for a comfortable stay. The current market trends show that user reviews are playing a vital role to select a hotel based on a safe location. Our work tries to find and recommend the dimensions that contribute more towards higher satisfaction levels of users in Asian continental top tourism city hotels listed by the Master and VISA Inc survey with item collaborative filtering and regression techniques based on Tripadvisor reviews from social media in recent times. The outcome of the work is helpful for Asian continental hotel management to concentrate more on the dimensions for obtaining better reviews. From an online travel platform perspective, this work suggests the contexts for the travel recommender system to gain things effectively from users.

Keywords: Tourism, Travel platforms, Hotel management, Hotel class, Travel purpose, Context, Collaborative, Recommender System.

1. Introduction.
Because of its influence and significance in e-commerce and the achievement of superior consumer acceptance, in the current days of growth, web technologies, the recommended systems (RS) have twisted the notice of business people and the common man to itself. Nowadays everyone wants to refresh themselves in the vacation by visiting the locations across the globe, mainly in the categories of middle and upper sections of users. At least once a year, these sections plan their vacations due to an increase in socioeconomic factors. Online Travel platform is one such opportunity for fulfilling their desires. Figure 1 shows the traditional recommender system architecture how it is working to fill fill customer needs.
Burke[7] talked about various matters. Types of algorithms suggested. It clarified there that Two methods of suggestion. Collaborative cooperation is first, Filtering (CF), and content-based filtering (CBF), respectively. The mixer of each of these algorithms is called hybrid filtering. The internet is a platform in which one can express the views and opinions among a community of users who are located globally[5]. The content generated by users through the internet encompasses diversity in different web domains and also in media[7]. Ratings and Reviews are a part of online content created by a user. The analyses and ratings provided by customers are one of the most prevalent and accessible information in the electronic World of Mouth[1,2]. In the vacation industry sector, eWOM helps users in decision making about the selection of a hotel to stay with the availability of online content. In the hospitality and tourism sectors, social media play a significant role[6][9].

The significant and widespread online activity in the tourism industry is the information searched by the travellers[10]. Many tourism studies indicate that the user relies on the information available in the online tourist platforms for planning their trips. In online travellers’ decision making, user-generated information leads to a vital role. User choices about travel plans are affected by the increase in the growth of online generated content. Before planning their trips, 50 per cent of travel planners depend on the content generated by previous users[8,20]. Opinion platforms are increasing in recent years, offering their users to create reviews and ratings online. TripAdvisor is one of the travel opinion platforms that allows users to express views and ratings regarding the hotel stayed experience.

Figure 1: A generic architecture of recommender system

2.Literature Survey

Prior studies on the recommender method discussed in the segment above. Previous research identifies similar factors Recommenders’ principles, such as the filtering of information and Recommendation algorithms used previously for the development of Recommendation of systems[20] which aid to know and to understand Comprehend the need of web-based recommenders in the modern age Technology.

2.1. Recommender Systems. Helping to end recommendation systems Users to help them locate goods and essential customer services that they are searching. Tan and He[13] tabled a necessary proposal Echo process, called resonance similitude (RES), a novel approach. Superior is the book resemblance. Predictive accuracy vs similar conventional Tourism means an overnight stay away from the regular place of residence. Tourism and hospitality sectors generate significant revenue in many country's economies. In recent years numerous e-commerce applications developed related to tourism and hospitality for attracting the users[12]. Current statistics indicate that majority of hotel room bookings related to travel are done online with a percentage of 71 by individual users[15]. Fasahte et al. Discussed if unrated products could be suggested and Using various filtering techniques, expected, and Using the Trip Advisor dataset, did tests. They also demonstrated a hybrid approach using ranking To predict user behaviour, data and textual content. Crespo Crespo Explain that Sem-Fit uses the experience of customers[16] point of view to apply the fuzzy logic technique to the relation between customers and hotels.

Hotels must concentrate on many vital aspects to obtain profitable revenues. One of the critical elements of the hotels is their presence online. Online reviews perform a significant role in knowing about the reputation of hotels, which act as an interface between users and providers of service offering useful information to make decisions about hotel stays[11][11]. To choose their accommodation in hotels, majority users depend on reputed online traveler platforms like Tripadvisor, Trivago, and Yelp, who provides the previous guest opinions based on social networks[15]. A recent study by Barclays indicates the potential growth of acquiring an additional 3.2 billion
dollars earnings by the hotel sector's income next decade if they pay more attention to online reviews those reviews to analyse and find out the best recommendations based on machine learning techniques[14,15].

Asia and the Pacific have been standing at the forefront of tourism development, having one-third of the world's economy. The hotel sector in this continent offers attractive tour packages in different hotel classes for users throughout the year. Hotel class and Trip-Type of a user take part in a crucial role in the selection of hotels to stay who have varied socioeconomic backgrounds[4,14]. Generalised Context-aware Recommender Systems is also one of the simple methods for acquiring views of previous users[3,15]. To build a customised hotel recommendation by Lin et al.[10], The techniques for text mining are used in combination with monitoring and browsing.

Rianthong et al.[13] It is concluded that the review ranking, prices and usefulness of the hotels must be taken on the upper side of the series in order to minimise the search expense of the customers. Using multiple regression analysis, a helpful model of stochastic programming was used. The most common and significant features for the classification of feelings are defined by Baghel and Lal[11] and grouped into seven classes, identified as basic features, seed word features, TF-IDF, punctuation-based features, phrase-based features, N-grams, and POS lexicons. Sharma et al., and Sharma, respectively.

Unfortunately, from the last three months in the entire world, the tourist roaming ratio falls due to the COVID-19 Pandemic[23]. Since most users rely on traveler domains in the selection of hotels they want to stay. To sustain the business and attract the customers continuously, one of the crucial factors taken into consideration by the traveler domains is users' past opinions for that hotel class and Trip-type along with the location. Users view changes from one location to another; many factors influence the satisfaction level, like hotel class and travel purpose[25]. This study tries to find the impact of trip-type on hotel class.[26] This research article tries to find the contextual segments which have a high impact on user overall rating from different hotel classes, trip-types in Asian continent hotels of top tourism cities, and recommends these contexts to both hotel management, traveler domains for future improvement[27].

3.Methodology:
The approach suggested uses the heterogeneous nature (textual and numerical) of data crawled in from social media. Data is collected from selected hotel websites (data sources) containing the keywords that are present in the active user search query. For the download of the requested data, a web crawler was used, and data for further processing was stored in the NoSQL Cassandra database. Numbers (such as votes, ranks, and a number of video views) and text (such as ratings and comments) typically contain E data. To obtain accurate recommendations, our system used ranks, votes, and feedback data to extract hotel features from it. In two simultaneous instances, the scheme works. Numeric ranks and hotel votes are normalised from each chosen source of data. Analysis info, on the other hand, is processed using the natural language processing lot for review mining and highlights are extracted in the form of a matrix of hotel functions. Furthermore, for those extracted functions, numerical polarity scores are computed using SentiWordNet, and the average polarity score is calculated. Weighted average polarity scores are now determined by aggregating the normalised rank score, voting score, and polarity score. Lastly, suggestions are co-recommendations. Regression approaches for machine learning techniques. To calculate the final score to find out the type of guest (solo, family, company, friends, couple, etc.) for the hotel, we have described a collaborative filtering other machine learning technique. The hotel recommendation method proposed is shown in Figure 2. In each of the five separate classes, the hotels' final recommendations are shown based on a specific guest category.

3.1. Feature Extraction Process.
The reviews are extracting from Tripadvisor, one of the leading e-tourism travel platforms. The data set consists of 37 cities across the Asian continent from different hotel classes having 15,977 reviews. From each city, we select 30 top hotels, at least with 200 to 1000 records. Maximum of reviews are coming from the last two recent years. This review contains the contextual dimensions Room, Location, Value, Cleanliness, Service, Sleep-Quality, and overall rating of the hotel. The item-item collaborative filtering recommendation algorithm used here replaces missing values of some contexts in this data set and backward elimination method in multiple regression algorithm which uses to obtain the significant impact of contextual dimensions on user overall satisfaction.
4. Results
After applying the O.L.S regression model to predict the significant contextual attributes on the overall user satisfaction level on different classes of hotels ranging from 1-5, The following screenshots show the results.

Table 4.1: List of cities in the Asian continent in consideration to extract reviews

| S.NO | CITY          | COUNTRY  |
|------|---------------|----------|
| 1    | Agra          | India    |
| 2    | Amman         | Jordon   |
| 3    | Bali          | Indonesia|
| 4    | Bangkok       | Thailand |
| 5    | Bangalore(Bengaluru) | India    |
| 6    | Beirut        | Lebanon  |
| 7    | Chennai       | India    |
| 8    | Cochin        | India    |
| 9    | Delhi         | India    |
| 10   | Denspar       | Indonesia|
| 11   | Dubai         | U.A.E    |
| 12   | Galle         | Srilanka |
| 13   | Giza          | Egypt    |
| 14   | Hongkong      | Hongkong |
| 15   | Hyderabad     | India    |
| 16   | Istanbul      | Turkey   |
| 17   | Jaipur        | India    |
| 18   | Japan         | Japan    |
| 19   | Johor Bahru   | Malaysia |
| 20   | Kuala Lumpur  | Malaysia |
| 21   | Katmandu      | Nepal    |
| 22   | Kuta          | Indonesia|
| 23   | Macau         | China    |
| 24   | Mecca         | Saudi Arabia |
| 25   | Medina        | Saudi Arabia |
| 26   | Mumbai        | India    |
| 27   | Osaka         | Japan    |
The above results indicate that all contextual segments make a significant contribution to the user gratification level in a hotel stay on several star hotels having in a range from 0.000 to 0.05, and Durbin-Watson is also in suggestable range for all hotel classes from 1.5 to 2.5. From the above observations of several hotel categories, we rank the contextual dimensions contribute more towards user happiness, also recommending these contexts to the recommender system to obtain valid recommendations is shown in the below table.

| Impact of Contextual Dimensions on Overall Rating in Several Star Hotels |
|------------------------------------------------------------------------|
| **Impact** | **Rank** | **Service** | **Value** | **Location** | **Cleanliness** |
| --- | --- | --- | --- | --- | --- |
| **service** | 1 | Service | Service | Service | Service |
| **value** | 2 | Value | Sleep-Quality | Sleep-Quality | Value |
| **location** | 3 | Location | Room | Value | Room |
| **cleanliness** | 4 | Cleanliness | Value | Room | Sleep-Quality |

Figure 7: Durbin-Watson is also in suggestable range for all hotel classes from 1.5 to 2.5.
Based on the rankings in above, all classes of hotels users are giving more importance to the context "Service" while in two, three-star hotels quality of the bed, i.e., Sleep-Quality, in one, four and five-star hotels amount spent for a room(Value) Have more impact in the user view. The context Cleanliness has an impact in one-star hotels, whereas least for the ambiance of a room(Room) since the user spent less amount comparing with other star hotels, so they give the least importance. The context Location does not make an impact on 3,4 and five-star hotels. Since users stay in hotels for various travel concerns, this study tries to explore the same context significance in several trip-types. Below are the results of the same trip-type in different star hotels.

Hotel class-2

Hotel class-3

Figure 8: Trip-Type: Family

Figure 9: Trip-Type: Couple

Figure 10: Trip-Type: Friends

Figure 11: Trip-Type: Solo

Figure 12: Trip-Type Friends

Figure 13: Trip-Type: N.A.

Figure 14: Trip-Type: Family

Figure 15: Trip-Type: Business
Table 4.3: Significant contextual segments in Trip-Type: Family

| Trip    | ** | *** | **** | ***** |
|---------|----|-----|------|-------|
| FAMILY  | 1. Service | 1. Service | 1. Service | 1. Service |
|         | 2. Sleep-Quality | 2. Value | 2. Value | 2. Value |
|         | 3. Cleanliness | 3. Sleep-Quality | 3. Room | 3. Cleanliness |
|         | 4. Location | 4. Location | 4. Location | 4. Room |
|         | 5. Value | 5. Room | 5. Cleanliness | 5. Location |
|         | 6. Cleanliness | 6. Sleep-Quality |

Table 4.4: Significant contextual segments in Trip-Type: Couple

| Trip     | ** | *** | **** | ***** |
|----------|----|-----|------|-------|
| Couple   | 1. Sleep-Quality | 1. Service | 1. Service | 1. Service |
|          | 2. Room | 2. Value | 2. Value | 2. Value |
|          | 3. Service | 3. Cleanliness | 3. Room | 3. Cleanliness |
|          | 4. Sleep-Quality | 4. Sleep-Quality | 4. Cleanliness |
|          | 5. Location | 5. Location | 5. Location |
|          | 6. Cleanliness |

Table 4.5: Significant contextual segments in Trip-Type: Business
| Trip     | **   | ***   | ****   | *****   |
|----------|------|-------|--------|---------|
| Business | 1. Service | 1. Service | 1. Service | 1. Service |
|          | 2. Sleep-Quality | 2. Value | 2. Value | 2. Value |
|          | 3. Room | 3. Sleep-Quality | 3. Sleep-Quality | 3. Sleep-Quality |
|          | 4. Location | 4. Cleanliness | 4. Cleanliness | 4. Cleanliness |
|          | 5. Cleanliness | 5. Room | 5. Room | 5. Room |
|          |        | 6. Location | 6. Location | 6. Location |

Table 4.6: Significant contextual segments in Trip-Type: Solo

| Trip     | **   | ***   | ****   | *****   |
|----------|------|-------|--------|---------|
| Friends  | Service | 1.Service | 1.Service | 1.Service |
|          | 1.Sleep-Quality | 1.Sleep-Quality | 1.Sleep-Quality | 1.Sleep-Quality |
|          | Room | Room | Room | Room |
|          | 2.Service | 2.Location | 2.Location | 2.Location |
|          | Value | Value | Value | Value |
|          | 3.Value | 3.Sleep-Quality | 3.Sleep-Quality | 3.Sleep-Quality |
|          | 4.Value | 4.Room | 4.Room | 4.Room |
|          | 5.Cleanliness | 5.Cleanliness | 5.Cleanliness | 5.Cleanliness |
|          | 6.Value | 6.Location | 6.Location | 6.Location |

Table 4.7: Significant contextual segments in Trip-Type: Friends

| Trip     | **   | ***   | ****   | *****   |
|----------|------|-------|--------|---------|
| N.A      | 1.Room | 1.Room | 1.Service | 1.Cleanliness |
|          | 2.Service | 2.Sleep-Quality | 2.Sleep-Quality | 2.Sleep-Quality |
|          | 3.Value | 3.Room | 3.Room | 3.Value |
|          | 4.Location | 4.Cleanliness | 4.Cleanliness | 4.Cleanliness |
|          | 5.Service | 5.Location | 5.Location | 5.Location |

Table 4.8: Significant contextual segments in Trip-Type: N.A (Not Answer)

Table 4.9: Impact Trip-Types on Hotel classes
Tables from 4.3 to 4.8 reflect the significant contextual segments in several travel purposes. From the above observations, Trip-Type makes a considerable impact on contextual segments in hotel classes. Contexts priority changes from one Trip-Type to another also its impact on user overall gratification. According to the results in table 4.9 in two-star hotels, Couple, four, five-star hotels Business and N.A trip-types shows a significant impact on user overall satisfaction.

5. Conclusion
In e-Tourism, the role of online travel platforms is a significant one that offers services and makes recommendations to their users for a comfortable stay in hotels upon prior views. The type of recommendations made by travel platforms is also vital in the hotel management eye in a business concern. To reach user dynamic satisfaction levels, every travel recommender system periodically broadly assesses the users’ views identify the things influencing more towards the expectation level, and makes necessary changes in the recommendation model. In this paper, we analysed user opinions of top tourism destinations in the Asian continent identified the contexts and travel purposes that show more interest in the user view to opt high gratification levels in several classes of hotels. Our observations indicate that user preferences vary with hotel classes and travel purposes.

Future Extension
Users are traveling around the world for several reasons; their preferences vary from one continent to another regarding the hotel stays. In the future, we will analyse other continent user opinions.

References

[1] Filieri, R., and F. McLeay. 2014. EWOM and accomodation an analysis of the factors that influence travelers’ adoption of information from online reviews. J. Travel Res. 53 (1): 44-57.
[2] Gretzel, U. (2006). Consumer Generated Content: Trends and Implications for Branding. e- Review of Tourism Research (eRTR), 4 (3).
[3] Gretzel, U, Hyan-Yoo, K and Purifoy, M (2007) Online Travel Review Study: The role and impact of online travel reviews. Laboratory for Intelligent Systems in Tourism, College Station
[4] Inversini, A., and I. Masiero. 2014. Selling rooms online: the use of social media and online travel agents. Int. J. Hosp. Manag. 26 (2): 272-292.
[5] Pew Internet & American Life Project (2006b). Internet Activities. Accessed online (December 1,2006) at: http://www.pewinternet.Qr.i/trends/Internet Activities 7- 9.06.htm
[6] Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. Tourism Management, 31(2), 179–188. doi:10.1016/j.tourman.2009.02.016
[7] Aparna Puvvadi, Kishore, Polurie Venkata Vijay," An Efficient Medical Image Watermarking Technique in E-healthcare Application Using Hybridization of Compression and Cryptography "Algorithm,JOURNAL OF INTELLIGENT SYSTEMS, JAN-2018,10.1515/jjisys-2017-0266.
[8] Patel, Ashok Kumar; Chatterjee, Snehamoy; Gorai, Amit Kumar"Development of an expert system for iron ore classification"ARABIAN JOURNAL OF GEOSCIENCES,AUG-2018,10.1007/s12517-018-3733-x
[9] Putluri, Srinivasareddy; Rahman, Md Zia Ur; Fathima, Shaik Yasmeen,"Cloud-based adaptive exon prediction for DNA analysis",HEALTHCARE TECHNOLOGY LETTERS,FEB-2018,vol-5,issue 1.start page25end page30,DOI10.1049/htl.2017.0032
[10] K. V. Daya Sagard , P Sai Durga , G. Kavya , K Sri Sravya , K. Krishna Veni,"Mobile based home mechanisation framework using IoT for smart cities",International Journal of Engineering & Technology, 7 (2.7) (2018) 266-269.
[11] K Sai Prasanthi , K.V.Daya Sagard,"Survey on secure protocols for data sharing through edge of cloud assisted internet of things",International Journal of Engineering & Technology, 7 (2.7) (2018) 92-95.
[12] K. V. Daya Sagard, U. Abbulu,K Chaitanya Kumar Reddy,"Using Fuzzy Clustering Techniques in Pharmaceutical Industry to Find Expired Medicines", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018.
[13] K.V. Daya Sagard,K Shyam Krishna, G. Lalith Kumar, P. Surya Teja, G. Charless Babu,"A Method for finding threatened web sites through crime data mining and sentiment analysis",International Journal of Engineering & Technology, 7 (2.7) (2018) 62-65.
[14]. A.Yasaswini, K.V. DayaSagar , K.ShriVishnu,V.HariNandan, PVRD Prasadara Rao,”Automation of an IoT hub using artificial intelligence techniques”,International Journal of Engineering & Technology, 7 (2.7) (2018) 25-27.
[15]. K.V.Day Sagar,M.Rupesh Chowdary,S.Mahesh,”Smart Crop Monitoring and Farming sing Internet of Things with Cloud” Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018.
[16]. Rakesh shirsanth,Dr.K.V.Daya Sagar,”A Review of fine grained access control techniques”,International Journal of Engineering & Technology, 7 (2.7) (2018) 20-24.
[17]. K.V.Day Sagar,Dr.S.Narayana,Dr.K.RaghavaRao G.Bhavya Deepika,M.SaiKiran Reddy,Developing Smart Kitchen Inventory tracking using Internet of Things, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018.
[18]. K. V. Daya Sagar*, Akella Pavan Kumar, Goli Sai Ankush, Thota Harika, Madireddy Saranya and Dasaraju Hemanth,”Implementation of IoT based Railway Calamity Avoidance System using Cloud Computing Technology”,Indian Journal of Science and Technology, Vol 9(17), DOI: 10.17485/ijsst/2016/v9i17/93020, May 2016, ISSN : 0974-6846.