

Extracting Features for Opinion Mining using Pointwise Mutual Information

N. Padmaja^{1#} T. Sudha^{2##}

[#]Research Scholar, Department of Computer Science and Engineering, School of Engineering & Technology,

Sri Padmavati Mahila Visvavidyalayam, 517502, Andhra Pradesh, India

^{##}Professor, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, 517502,

Andhra Pradesh, India

¹gowripadma99@gmail.com

² thatimakula_sudha@yahoo.com

Article Info Volume 81 Page Number: 6098 - 6118 Publication Issue: November-December 2019

Abstract:

Opinion mining has been developing rapidly in the past years mainly because of its huge set of applications and the scientific barriers it poses. The opinions of tourists are useful for the stakekolders and for public when making some decisions. Opinion Mining is a way to extract data through search engines, social networks and web blogs. The essential part in opinion mining is to analyze and retrierve the feedback of tourists to disover their opinions on different hotels. Now users can actively use IT to search others opinions based on tourism. This study proposes point wise mutual information for summarizing and extracting opinions denoted by tourists in tourism associated internet platforms. The pointwise mutual information is used to decide the semantic orientation of opinions by providing queries to a search engine where poor and excellent words have been considered as a boundary for negative and positive inference words. In this study the PointWise Mutual Information has been used to identify the adjectives and noun phrases related with the target. Retrieving opinions from user generated reviews regarding the perspectives specific to hotel services are helpful for both the clients who are viewing for accomodation and also hotels attempting to develop their services. The proposed system retrieves the hotel reviews from online and categorize them using opinion mining techique. This study explains the importance of opinion mining using pointwise mutual information in the growth of tourism. This study focuses mainly on the opinions of travel behavior of tourists. The datasets collect reviews from the following cities namely Chicago, Beijing, New York, Dubai, San Francisco, London, New Delhi, Las Vegas and Shanghai.

Article History Article Received: 5 March 2019 Revised: 18 May 2019 Accepted: 24 September 2019 Publication: 28 December 2019

Keywords: Opinion Mining, PMI, Hotel Reviews, Opinions, Tourist travel behaviour

1. Introduction:

The awareness in opinion mining has been developing rapidly in the past years mainly because of its huge number of applications and the scientific barrier it poses. Opinion mining is an essential research area because of its applications in different fields. The reviews of users are useful for



stakeholders and for public when making some decisions. Opinion mining is a way to extract data through web blogs, social networks and search engines (Kavitha et al, 2017; Khan, 2014). Alaei et al (2019) has mentioned that the growth in technology have changed fundamentally how data is consumed and generated by marketers in the field of tourism. Tourists can access various data sources and they can produce their own concept and share their experiences and views. The concept of tourism shared through social media has become an impactful source of data that influences tourism in terms of both performance and reputation. However the data volume on internet has met a level that makes manual processing impossible requesting new analytical methods. Sentiment analysis or opinion mining is developing rapidly as an automated method of investigating semantic meaning and relationships in reviews. Neidhardt et al (2017) has mentioned that the tourism sector is a sector where the experience of traveller is crucial for its reputation and growth and has adapted to developing technique mainly and the feasibility of new sources of data. Many services of tourist are now feasible internet through online travelling sites. Travel is one of the major concepts on social media for Twitter and Facebook. instance on Tourism has been identified as the major field in online engagement term. Wang (2016) has stated that tourism depends on feedback and emotions of positive customer and the visitor satisfaction concept is of huge essential. Changes in technology associated to internet involving tablets and smart phones have revolutionize the tourism sector from a

person to person service and brick and mortar sector to a heavily digitally assisted and omnipresent network of travel service. Individual groups or travellers have much control over building, personalizing and planning their trips. They not only communicate with a set of online intermediaries and online platforms to expand their knowledge associated to decision making and travelling in tourism but also relate with other travellers who share their reviews. Travellers have access to online sector to offer feedback and make suggestions for other travellers (Yang et al, 2017). According to Afzaal et al (2016) tourism is a dynamically developing sector and essential for several nations and regions as major industry. Thousands and hundreds of tourists visit the tourist spots each year and share their reviews on different sites. These opinions offer an overall review of an opinion holder considering the tourist spot. However, there are several sets of opinions which are feasible on a specific spot and it is critical for a general user to read or review all the feasible opinions and determine on whether to visit a spot or not. Several methods of opinion mining can be used to deal with huge set of opinions and these methods helps to categorize the opinions into negative and positive.(Ravi and Ravi, 2015, Gowtamreddy, 2014).

Zhang and Liu (2016) have stated that with the rapid development of social media sentiment analysis or opinion mining has become one of the most active areas of research in NLP extraction and detection of opinions from online reviews is a part of new research area evolved in the past decade. Opinion mining also referred as sentiment analysis studies the



classification and determination of feelings or opinions denoted in text through the computing machines use (Songpan et al. 2017). The barrier of research area is to retrieve knowledge from unstructured information. The reviews comprise of opinions denoted in natural language similar to people but uninterrupted by personal computers (Bucur, 2014). Aurchana et al (2014) has mentioned that opinion mining analyse the feelings that is opinions, attitudes and emotions behind the words using NLP tools. It views beyond the number of comments, likes or shares which the user acquires on a promotion campaign, release of product, blog post and video to perceive how users are answering to it. Opinion holder, opinion target and opinion are used to retrieve opinions from various sources of online. An opinion can be denoted in two ways comparative opinion and direct opinion. All opinions are stored in a document. The opinion mining major tasks are: 1) aspect identification; 2) choosing sentiment essential aspects; 3) classification into negative and positive aspects; and 4) results summarization (Lal and Asnani, 2014). Tourism marketers uses two kinds of internet content for their opinion analysis namely reviews of tourism acquired from social media posts (for example Twitter) and professional sites (for example, TripAdvisor, Ctrip). Both kinds of sources generally comprise of small text. For example, Twitter permits tweets of around 140 characters in length lending them to sentence level opinion mining. Automatic and Manual annotation methods were employed to mention the opinions to evaluate and train the methods of sentiment analysis (Rossetti et al, 2015).

Dadhaniya and Dhamecha (2017) have mentioned that opinion mining can be helpful in many ways. It can support markers estimate the success of an advertisement promotion or new launch of product decide which service or product versions are familiar and recognize which demographics dislike or like specific features of product. Situmorang et al (2019) has stated that sentiment analysis in tourism has an essential effect on tourism sentiment growth because analysis analyses and identifies emotions and opinions comprising reviews provided by tourist's service provider or tourists (Gao et al, 2015).

Georgieva-Trifonova et al (2018) has mentioned that the pointwise mutual information offers scoring n-grams by comparison of the multiword expression candidate frequency to the multiword expression components frequency. With this measure the bi-gram multiword expression identification is performed. Fatima and Arora (2016) have stated that the point wise mutual information method is calculated by estimating the log of frequency ratio of two words that exists together to separate frequencies of 2 words. Semantic orientation is decided by estimating the difference of point wise mutual information values evaluated against negative words from pointwise mutual information values evaluated against positive words. Qin (2016) has that the Point wise mutual stated information is also the measure of the semantic relation strength between two words which compares the observing word1 and word2 probability together with the observing word 1 and word 2 probabilities independently. The PMI



between two words is calculated using the below given formula:

Point - Wise Mutual Information (W_1, W_2) = $log_2[\frac{Pb(W_1 \& W_2)}{Pb(W_1) Pb(W_2)}]$

Where W_1, W_2 indicates word 1 and word 2, Pb $(W_1 \& W_2)$ denotes the two words probability existing together If there is a genuine association between 2 words then Pointwise mutual information (word1, word2)>0 denotes that the co-occurrence of two words probability is greater. If the Pointwise mutual information (word1, word2)<=0 then there is no essential association between two words or they are in balancing distribution. Gomez et al (2018) has stated that the point wise mutual information is used in the tourism sector by collecting the general opinions of user about the travelling experience, services of a hotel where there are keywords that indicates the tourist emotions even though various emotions are shown in the opinions of tourists. The point wise mutual information permits classification of text based on negative or positive polarity. This method acquires a phrase of text and compares it with feasible database on internet to decide an opinion method that tries to categorize the text polarity. Thus it can be inferred that opinion mining is helpful for classifying and retrieving reviews posted by users on social media sites easily and also develops the tourism growth.

2. Literature Review:

In the study of Mankar and Ingle (2014) sentiment analysis or opinion mining is the study of computation of emotions or opinions towards things or aspects. The perspectives are nothing but the components or attributes of events, individuals, products, organizations and topics. Opinion mining has been an active area of research in NLP (Natural Language Processing) and web mining presently. With the huge development of electronic domain there are huge number of product choices feasible and people tend to review others viewpoints before purchasing a product. An opinion mining based on aspect method helps in examining views about product attributes and features. This study is based on retrieving aspects and associated sentiments of consumers on the domain of tourism. This provides a method to discover preferences of customer about the services and products of tourism using statistical opinion mining. The proposed system attempts to retrieve both implicit aspects as well as explicit aspects from reviews of customers. Thus, it raises the opinions sentiment orientation. Many studies were based on customer's explicit opinions. This system attempts to extract implicit opinions. This study proposes point wise mutual information to discover the emotions or implicit aspects hidden in the opinions. Due to the developing feasibility of unstructured reviews the proposed system offers an overview form of data that is acquired from the opinions to enhance customers with crisp or proper results.

The study of Bucur (2015) proposes a platform for summary and retrieval of reviews denoted by customers in tourism associated online platforms. Retrieving reviews from user generated opinions, considering the aspects particular to services of hotel are helpful both to clients



viewing for accommodation and also hotels attempting to develop their services. The proposed system retrieves hotel opinions from online and categorizes them using a technique of opinion mining. Platform is estimated using a manually pre-categorized set of data of user opinions. In this study the algorithms efficiency is examined using the domain of text mining particular measures and are proposed approaches for developing the results.

Sharma et al (2015) has mentioned that due to the rapid development in tourism, tourists like to predict data regarding various places. With a rapid development in web usually tourists share their social site websites. These websites have become the major source of data for tourism but due to huge information and mixed emotions it becomes critical for a tourist to make a decision regarding the travelling. For this issue this study proposed the tour view system a tool that provides a group of summarization methods to support users digest the wide availability of reviews in an easy way. The main aim behind this system is that tourists can easily retrieve useful and subjective data. This system decreases the time needed for browsing and easily assists development of tourism.

Ma et al (2016) has mentioned in their research that the satisfaction of tourists has become an essential indicator of tourism growth. Sentiment analysis on hot discussion on travel websites and data of comments and Weibo can judge real time satisfaction trend of tourists with scenic places according to the sentiment tendency intensity. Regarding the deficiencies in present sentiment analysis this study uses the strength value and polarity value to estimate the sentimental words sentimental intensity. HIT_CIR TongyiciCillin is used to extend the sentimental words synonyms in order to decrease the effect of words not which is not in HowNet and certain words of sentiment with reduced frequency in corpus. Then the traditional semantic similarity method based on HowNet is improved according to sentiment analysis features integrating it with the approach based on point wise mutual information and relations of syntactic dependency. Greater accuracy is displayed by the results of experiment.

Parashar Sharma (2016)and has mentioned in their study that online media become one of has the strongest information sources nowadays. People visit several locations and share their views on the tourism sites. Tourism sites is a most easy and capable way of data. Opinion mining is a method to retrieve knowledge from reviews generated by user. Naïve Bayes is a probabilistic classification method and Support Vector Machine is a classifier that categorize into 2 classes with huge margin. This study proposes a hybrid approach integrates SVM and Animal migration optimization in an optimized manner and the measures of performance namely recall, precision and accuracy estimated for are justification.

Dave et al (2017) has stated that in nowadays trend of social media everyone is utilizing online for each perspective of their lives whether it is purchasing, interfacing with each other or getting data. The main aim of this research is to enclose one another but essential perspective is



tourism. Since tourism plays an essential part in the economy of any nation it is essential to discuss on this perspective. The experience of traveller is consequential in any business so as to develop the experience of customer, different places reviews, restaurants and reviews considered and after are examining them an overall opinion of that spot is produced so the traveller could determine which spot to visit and which not, so this saves both money and time and experience of traveller is also positive,

Gomez et al (2018) has mentioned in their research that one of the information sources to predict the customer satisfaction level is the group of reviews that is viewed on sites namely Trip Advisor. An issue emerges when there is huge set of administrators and reviews have difficulty in recognizing those that indicate the feelings of users from the reviews. For this research certain applications produced in the programming language of Python were utilized to allot a bad and outstanding criterion by adding a textual feeling. The classifier of SentiStrength was used which examines text to return a value of negative or positive sentiment and these two contexts were used statistically and mathematically to acquire referential information based on the examination of ROC curve with values of true negatives and true positives as well as false negatives and false positives. This study proposed a methodology based on Point to acquire Mutual information the semantic phrases orientation that make up a review.

In the study of Gupta et al (2018) websites like Trip Advisor has become an essential part of human being. It provides different reviews about the hotels and supports them to determine which hotels to visit or stay in. The reviews of travellers play an essential part in impacting others. Hence these reviews become very essential in handling the hotel quality. The main aim of the study is to gather all reviews from web and produce rating based on it with the use of classification of data mining that is decision tree. The C4.5 method of decision tree is used for the above purpose using Tanagra ML tool to produce the model of data. The benefit of using decision tree algorithm is that the rule set can be generated easily and by examining every tree level which develops the specific quality of service.

Hermanto et al (2018) has mentioned in their study that the present destination of tourism has developed widely which is influenced by the role of technique that has a major impact on freedom of access in seeking data. A tweet has the ability of comprising data or condition about the tourist places that they have or will visit such as the visitors experience in the tour, opinion of visitors on the tourist place and other tourist attraction places. The Naïve Bayes Method uses training information to create the probability of every criterion for various classes nevertheless the criteria's probability value can be developed to generate opinion mining towards the reviews on twitter. This study has been carried out by categorizing an opinion in comments forms in two classes which is negative and positive with the accuracy level that is impacted by the process of training Based on that it can be inferred that the public sentiment data to the tourist attractions involved in positive sentiment.



Rai and Ahirwal (2018) has stated in their research that in present years for analysis of text sentiment, lexicon-based methods has been utilized essentially. Several research studies have been utilized lexicon approach for analysis of review sentiment. Α new review sentiment lexicon generation approach is proposed using the POS tags in the lexicon. Training the review information is chosen from an unlabelled corpus based on their word ranks as assessed by SentiWordNet. Consequently, sentiment a analysis structure is suggested based on a whole lexicon that utilizes the machine learning method to examine the class of reviews that is both negative and positive. For analysis of simulation a dataset of tourism reviews is prepared by gathering varied reviews from various sites. The proposed algorithm is examined using three classifiers that is K-nearest neighbour, random forest and support vector machine. After the results of simulation K-NN provides nearly 80 percent accuracy, random forest provides 93 percent accuracy and SVM provides 89 percent accuracy. The results reveal that the structure of sentiment analysis based on the proposed lexicon is efficient for records of tourism that utilize a classifier of random forest.

In the study of Rajan and Shyam (2018) tourism is a rapidly developing travelling practice across national and international borders to acquire different objectives. Tourism is developing to be an essential income source across the world. The reviews of online customers on tourism destined restaurants and hotels playan essential part in decision making while selecting destination restaurants and hotels. Text reviews and travel blogs in travel site plays an essential part while selecting hotel as well as destination for a given place. Subsequently tourists utilize similar data to match their choices. Similarly, the authority of hotel employs the same data to develop the service of customers. A much effortless and better inference could be employed if entire reviews are examined for a given place rather reading sample reviews. Scraping reviews of online customer from travel sites and opinion mining on reviews of hotel by hotel authorities as well as customers could be used to develop the service to be provided by the executives of hotel through meticulous decision making. The classification of sentiment is a supervised learning carried out on text data to mine sentiment of a specific keyword existing in the text.

Sindhuja et al (2019) proposed a study on an Enhanced Sentiment Analysis Model for Incorporating Implicit Aspects. An aspect-based sentiment analysis method supports in examining opinions about product attributes and features. An aspect level analysis of sentiment would define to the sentiment related with product aspects. The aspects are about the components or attributes of products, organizations and topics. This study is based on retrieving and associated opinions aspects of customers. This provides a method to discover the customer choices about the services using statistical opinion mining. The proposed system retrieves both the implicit as well as explicit aspects from customer opinions. First it performs certain pre-processing steps to remove unneeded words in the sentences review. Second it carries out the extraction of



semantic aspects and then it shows the review status. This system retrieves the implicit aspects using rule-based classifiers. Due to the development of unstructured opinions the proposed system provides an overview of the data that is acquired from the reviews to enhance users with appropriate results. The existing system of this study applies point wise mutual information-based analysis of semantic association and co-occurrence association rule mining to recognize implicit features.

The below Figure shows the reviews of the opinion mining using point wise mutual information in tourism sector:

S.No.	Author	Year	Technique used	Benefits of the technique
1	Mankar and Ingle	2014	Statistical Opinion	Retrieve implicit
			Mining	opinions, provides an
				overview of the data that
				is acquired from opinions
				to enhance customers
				crisp or pinpoint
				outcomes
2	Bucur	2015	Opinion Mining	Retrieves and categorizes
			technique	opinions of hotel posted
				by users on tourism sites
3	Sharma et al	2015	TourView	Provides a group of
				summarization methods
				to support users digest
				the wide availability of
				reviews in an easy way
4	Ma et al	2016	HIT-CIR TongyiciCilin	Extend the sentimental
				words synonyms to
				reduce the effect of
				words with reduced
				frequency in corpus
5	Parashar and Sharma	2016	Animal Migration	Optimize doubt opinions
			Optimization and	and categorize into
			Support Vector	negative and positive
			Machine Technique	opinions on fitness value



				basis.
6	Dave et al	2017	Machine Learning	Saves time and money
			Approach	and the experience of
				customer is positive
7	Gupta et al	2018	C4.5 Decision Tree	Rule set can be generated
			Classification	easily and service quality
				can be improved by
				examining each tree level
8	Gomez et al	2018	Feelings Mining	Decide the satisfaction
				level of service of hotels
				through their reviews and
				recognize the text phrases
				polarity in these online
				reviews
9	Hermanto et al	2018	Naïve Bayes	Generates the probability
			Classifier	of every criterion for
				various classes and the
				criteria's probability
				value are optimized to
				produce opinion mining
				on twitter reviews on
				tourist places based on
				the proves of
				classification that has
				been performed
				consequently.
10	Rai and Ahirwal	2018	Lexicon Based	Improves the reviews
			Sentiment Analysis	recognition performance
			Structure	
11	Rajan and Shyam	2018	Text Mining Method	Removes the excess time
				used while taking reviews
				of online customers from



				tourism sites and
				automate scraping of
				online text review from
				numerous tourism sites
				which would offer exact
				and better insights.
12	Sindhuja et al	2019	Statistical opinion	Extracts the implicit as
			mining and rule	well as explicit aspects
			based classifiers.	from customer opinions
				and remove unneeded
				words in sentences
				reviews.

Table 1: Reviews of the opinion mining using point wise mutual information in tourism

sector

Source: Author

3.Design of the System:

3.1 Proposed System:

This section describes Extracting Features for Opinion Mining using Pointwise Mutual Information. The data is obtained from the opinions provided by tourist travellers in the social media sites. The opinions of tourist travellers have been analysed to identify whether the tour has provided them the satisfaction or not and also analyses to what extent the opinions of female tourists differ from that of male tourists travellers. The below figure the proposed system flow diagram:





Figure 1: Proposed System Flow Diagram

Source: Author

From the above figure the following steps are used:

Step 1: The first step is to collect the data of tourist reviews used in online sites

Step 2: Following the tourist reviews the POS (Point of scale) tagging is applied to extract and identify the two phrases of

word that comprises the <Noun, Adjective> that are adjacent to each other.

Step 3: Then only those features are kept visualizing together at least three times in the reviews.

Step 4: Then the occurrences frequency in the reviews for the complete set of



retrieved nouns is computed and the frequency of those features which is larger than 0.02.

Step 5:Finally, the PMI is evaluated using the below formula:

$$PMI (A, N) = \log P(NounAdj) - \log P (Adj)$$

Where P(NounAdj) is the probability where the Adj adjective exists given the Noun data. This study prunes having the point wise mutual information is less than zero.

3.2 Description of Dataset:

The data set used in this study comprises 259000 reviews from TripAdvisor and 42230 reviews from Edmunds. The dataset comprises hotel reviews in nine different cities namely Beijing, Dubai, London, San Francisco, New Delhi, Montreal, Chicago, Las Vegas and Shanghai. There are nearly 80 to 700 hotels in every city. The retrieved field involves the date, title of review and the complete review (Ganesan and Zhai, 2012).

4. Discussion and Results:

4.1 Discussion:

This study discusses the hotel reviews of tourist travellers from TripAdvisor and Edmunds sites. The reviews are based on positive, negative and neutral opinions of tourist travellers based on their experience in hotels. The reviews are posted with the name of the tourist, title of their review and the positive and negative opinions about the hotels. These hotels reviews will be useful for tourist travellers to have information about which hotels is best in which city. There is several numbers of hotels in all the nine cities mentioned in this study. The tourist traveller can view all the opinions of hotels and compare which hotel is the best to book. The higher to lower ratings are given for all the hotels based on the opinions of tourist travellers.

4.2 Results of the Proposed System:

The results of this study will be the count of features that persuade tourist towards each cities respectively. This study will focus on extractingFeatures for Opinion Mining using Pointwise Mutual Information based on the reviews and ratings made by the tourist travellers in social media sites. The charts with count of features in X-axis and Top 10 features in Y-axis is depicted for all cities below.

4.2.1 Beijing:





Figure 2: Count of Features of Beijing

with good location and lesser number of features with nice hotel.

It can be inferred from the above chart that Beijing has greater number of features **4.2.2 Chicago:**



Figure3: Count of Features of Chicago



It can be understood from the above chart that Chicago has greater number of features with great location and lesser number of features with nice hotel.

4.2.3 Dubai:



Figure4: Count of Features of Dubai

Inference:

It can be understood that Dubai has greater number of features with good value and lesser number of features with fresh fruits.





Figure 5: Count of Features of Las Vegas



It was found from the above figure that Las Vegas has greater number of features with first time visit of tourists and lesser number of features with great view.

4.2.5 London:



Figure6: Count of Features of London

Inference:

It can be inferred from the above chart that London has greater number of features with great location and lesser number of features with next door.

4.2.6 Montreal:



Figure7: Count of Features of Montreal



It can be understood from the above figure that Montreal has greater number of features with great location and lesser number of features with great place.

4.2.7 New Delhi:



Figure8: Count of Features of New Delhi

Inference:

It can be found from the above chart that New Delhi has greater number of features for hot water whereas lesser number of features for best hotel.



4.2.8 San Francisco:

Figure 9: Count of Features of San Francisco



It can be understood from the above graph that San Francisco has greater number of features with great location and lesser number of features for friendly staff.

4.2.9 Shanghai:



Figure10: Count of Features of Shanghai

Inference:

It was found from the above graph that Shanghai has greater number of features for good hotel and lesser number of features for excellent hotel.

4.3 Comparison of the Results of the

The common features used in the hotels for the nine cities are Good hotel, Good location, good value, next door, great location. Each feature of nine cities is compared and the results are depicted below:



Figure 2: Count of Features of Beijing



It can be inferred from the above graph that compared to all the cities London has all common count of features with Good hotel, Good Location, Good Value, Next Door and Great Location.

5. Conclusion:

This study mainly proposes a new approach by extracting features for Opinion mining using point wise mutual information. Opinion mining has become familiar in the past few decades by the raise of online shopping and ever developing number of user generated contents involving reviews of hotel. Opinion mining using social media information is a very essential research area due to its efficiency and usefulness because huge amount of text can be examined in a much reduced amount of time with better accuracy which is not feasible manually. This offers useful data for an online sector which utilizes to offer internet users with useful data when making a booking. Opinion mining is the method of categorizing a given set of text into three polarity classes that is negative, neutral and positive. The right of expression sentiments and opinions about everything on social media sites enhances huge number of data each moment in the form of blogs, tweets, forums and news. By using this information the data can be analysed and mined for different prediction and decision making purposes. This study aims at predicting tourism associated opinionated phrases and recognizing their attraction targets of Opinion mining based on tourism. features is useful for the tourist travellers as well as for tourist guide. Tourist travellers predict it advantageous to view

the opinions before traveling anywhere in the site whereas tourist guide use this data to verify how their tourism information is being received by travellers and they can also verify with their rivalry tourist sites. uses pointwise This study mutual information for retrieving the features and their respective opinions as well as the barriers that exist while mining the opinions. It can be concluded that in tourism domain the feature based analysis develops the platform performance and due to numerous perspectives tourist travellers express opinions about and combined opinions that occurs in the reviews. In future, real time opinion mining system while intrusion of data is considered randomly.

References:

- Afzaal, M., Usman, M., Fong, A. C. M., Fong, S., &Zhuang, Y. (2016). Fuzzy aspect based opinion classification system for mining tourist reviews. *Advances in Fuzzy Systems*, 2.
- Alaei, A. R., Becken, S., &Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), pp 175-191.
- Aurchana P, Iyyappan R and Periyasamy P (2014), Sentiment Analysis in Tourism, International Journal of Innovative Science, Engineering and Technology, Volume 1, Issue 9, pp 443-450.
- Bucur, C. (2014). Opinion Mining platform for Intelligence in business. *Economic Insights-Trends and Challenges*, 3(3), pp 99-108.
- 5. Bucur, C. (2015). Using opinion mining techniques in



tourism. *Procedia economics and finance*, 23, pp 1666-1673.

- Dadhaniya B and Dhamecha M (2017), A Study on Feature extraction and summarization using Machine Learning and Opinion Mining, International Journal of Engineering Development and Research, Volume 5, Issue 1, pp 384-387.
- Dave V, Shah D, Suthar D, Prajapati B and Puvar P (2017), "Sentiment Analysis of Tourists Opinions of Amusement, Historical and Pilgrimage Places: A Machine Learning Approach". International Journal of Computer Trends and Technology (IJCTT) V46(2):51-55.
- Fatima J and Arora D (2016), Feature Based Opinion Mining: A Review, International Journal of Innovative Research in Science, Engineering and Technology, Volume 5, Issue 5, pp 7345-7351.
- 9. Ganesan, K., &Zhai, C. (2012).Opinion-based entity ranking. *Information retrieval*, *15*(2), 116-150.
- 10. Gao, S., Hao, J., & Fu, Y. (2015, June). The application and comparison of web services for sentiment analysis in tourism. In 2015 12th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1-6).
- Georgieva-Trifonova, T. L., Stefanova, M. E., &Kalchev, S. D. (2018). Customer Feedback Text Analysis for Online Stores Reviews in Bulgarian. *IAENG International Journal of Computer Science*, 45(4).
- 12. Gomes, H. F., Lozada, E. F., Martínez, C. E., Baño, F. P., Álvarez,

G. A., Culque, W. V., Bustamante-Sánchez, N. S. and Cevallos, R. E. S. (2018). Introduction to the Analysis of Tourism Services through Feelings Mining.Journal of Information Systems Engineering & Management, 3(4), 33.

- 13. Gomez, A. H. F., Bano, N. F. P., Martinez, C. C. E., Gómez, G. A. Á., Culque, T. W. V., & Bustamante-Sanchez, N. S. (2018,June). Introduction of the opinions of tourist services based on sentiment. In 2018 13th Iberian Conference on Information **Systems** and Technologies (CISTI) (pp. 1-8).
- 14. Gowtamreddy, P. (2014). *Opinion mining of online customer reviews* (Doctoral dissertation).
- 15. Gupta S, Jain S, Gupta S, Shruti and Chauhan A (2018), Opinion Mining for Hotel Rating Through Reviews Using Decisio9n Tree Classification Method, International Journal of Advanced Research in Computer Science, Volume 9, No. 2, pp 180-184.
- 16. Hermanto, D. T., Ziaurrahman, M., Bianto, M. A., &Setyanto, A. (2018), Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier. In *Journal of Physics: Conference Series*, IOP Publishing, Vol. 1140, No. 1, p. 012037
- 17. Kavitha S, Sathyavathi S, Prabhakaran S, Swathi S and Rajkumar R (2017), Opinion Mining On Tourism, International Journal of Innovative Research in Science and Technology, Volume 3, Issue 8, pp 128-131.



- Khan, K., Baharudin, B., Khan, A., &Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. Journal of King Saud University-Computer and Information Sciences, 26(3), 258-275.
- 19. Lal, M., &Asnani, K. (2014). Implicit aspect identification techniques for mining opinions: a survey. *International Journal of Computer Applications*, 98(4).
- 20. Ma, Z., Du, J., & Zhou, Y. (2016). Sentiment analysis based on evaluation of tourist attractions. In *Proceedings of the 2015 Chinese Intelligent Systems Conference* (pp. 375-382).Springer, Berlin, Heidelberg.
- 21. Mankar, S.A and Ingle, M. (2015). Implicit sentiment identification using aspect based opinion mining. International Journal on Recent and Innovation Trends in Computing and Communication, 3(2015), pp 2184-2188.
- 22. Neidhardt, J., Rümmele, N., &Werthner, H. (2017).Predicting happiness: user interactions and sentiment analysis in an online travel forum. *Information Technology & Tourism*, 17(1), pp 101-119.
- 23. Parashar, P and Sharma, S (2016), Opinion Mining of Tourism Review Using Hybrid Technique of Support Vector Machine and Animal Migration Optimization, International Journal of Innovations in Engineering and Technology, Volume 7, Issue 1, pp 718-731.
- 24. Qin Z (2016), A Framework and practical implementation for

sentiment analysis and aspect exploration, University of Manchester Thesis.

- 25. Rai, P., &Ahirwal, R (2018), Tourism Review Sentiment Analysis using Lexicon Features and Machine Learning Approach, Volume 5, Issue 4, pp 46-54.
- 26. Rajan, A., &Shyam, A. V. (2018). Sentiment Analysis on Customer Reviews in Tourism–A Text Mining Approach. Available at https://www.amrita.edu/sites/default/f iles/sentiment-analysis-on-customerreviews-in-tourism-a-text-miningapproach.pdf, accessed on 11th October 2019.
- 27. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- 28. Rossetti, M., Stella, F., &Zanker, M. (2016). Analyzing user reviews in tourism with topic models. *Information Technology & Tourism*, 16(1), pp 5-21.
- 29. Sharma, D., Kulshreshtha, A., &Paygude, P. (2015).Tourview: Sentiment based analysis on tourist domain. *International Journal of Computer Science and Information Technologies*, 6(3), pp 2318-2320.
- 30. Sindhuja R and Nandhini M D S (2019), An Enhanced Sentiment Analysis Model for Incorporating Implicit Aspects, International Journal of Innovative Research in Science, Engineering and Technology, Volume 8, Issue 2, pp 855-861.
- 31. Situmorang, V., Elyani, T., Tambunan, R., &Gulto, Y. (2019),



Applying Opinion Mining Technique on Tourism Study Case: Lake Toba. In *Journal of Physics: Conference Series* (Vol. 1175, No. 1, p. 012091). IOP Publishing.

- 32. Songpan, W. (2017), The analysis and prediction of customer review rating using opinion mining. In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 71-77).
- 33. Wang, Y. (2016). More important than ever: Measuring tourist satisfaction. Griffith Institute for Tourism, Griffith University.
- 34. Yang, Y., Mao, Z., & Tang, J. (2018).Understanding guest satisfaction with urban hotel location. *Journal of Travel Research*, 57(2), pp 243-259.
- 35. Zhang L and Liu B (2016), Sentiment Analysis and Opinion Mining, Available at <u>https://www.researchgate.net/publicat</u> <u>ion/305622270_Sentiment_Analysis_</u> <u>and_Opinion_Mining</u>, accessed on 12th October 2019.