Is Sentiment Analysis in Social Media Influence the Business Growth of Tamilnadu Tourism?

1 author:

Senthil Veerasamy

Thiagarajar School of Management

8 PUBLICATIONS **24** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   IT Management View project

Is Sentiment Analysis in Social Media Influence the Business Growth of Tamilnadu Tourism?

V. Senthil

Abstract--- *This paper is an exploratory study of how Social Media Sentiments influence the business growth of Tamilnadu Tourism. Sentiment Analysis deals with the analysis of emotions, opinions, facts and sentiments in the sentences which are expressed by the people. The objectives of this paper are: First, social media interactions on the state tourism web portals managed by the Destination Management Organizations (DMOs) are gathered and analyzed. Secondly this paper presents an approach to extract sentiments associated with a phrase or sentence, to track attitudes and feelings of the people by analyzing blogs, comments, reviews and tweets. The process of filtering irrelevant and unhelpful reviews and quantification of sentiments are analyzed using Natural Language Processing techniques. The social media interactions such as Likes, Talking about this, Tweets, Followers, Following, Video views and Photo views are analyzed in this study, to explore how these interactions are influencing the growth of Indian and Tamilnadu Tourism business.*

Keywords--- *Social Media, Digital Sentiment Analysis*

I.   INTRODUCTION

INFORMATION Technology infrastructure and services are integral part of any business and is unavoidable in today's competitive business scenario. The Proliferation of social media, cloud computing and mobile technology has enhanced the quality and quantity of data generated every day. Internet of Things (IOT), Internet of Vehicles (IOV) which includes not only an interconnection of things but also an exploding digital network of people and data. During the last decade, the Business infrastructure has become digital with increased interconnections among products, processes and services. Across many firms spanning different industries and sectors, digital technologies (viewed as a combination of information, computing, communication and connectivity technologies) are fundamentally transforming business strategies, business processes, firm capabilities, products and services and key inter-firm relationships in extended business networks Anandhi Bharadwaj et al., [1]. These digital technologies are fundamentally reshaping traditional business strategy as modular, distributed, cross functional and global business processes that enable work to be carried out across boundaries of time, distance and function.

Opinions are central to almost all human activities and are key influencers of our behaviors. The beliefs and perceptions of reality, and the choices we make are to a considerable degree, tradition upon how others see and evaluate the world. For this reason, when need to make a decision human often seek out the opinions of others. This is not only true for individuals but also true for organizations. With the explosive growth of social media (e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites) on the Web, individuals and organizations are increasingly using the content in these media for decision making. Nowadays, if one wants to buy a consumer product, one is no longer limited to ask one's friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available. Extracting and summarizing the opinions in them is called as opinion mining.

In recent years, it have been witnessed that opinionated postings in social media have helped reshape businesses, away public sentiments and emotions, which have profoundly impacted on our social and political systems. Such postings have also mobilized masses for political changes such as those happened in some Arab countries in 2011. It has thus become a necessity to collect and study opinions on the Web. Of course, opinionated documents not only exist on the Web (called external data), many organizations also have their internal data, e.g., customer feedback collected from emails and call centers or results from surveys conducted by the organizations. This paper involves the gathering of end user reviews about the Tamilnadu tourism from heterogeneous social sites including Facebook, YouTube, blogger and twitter and analyzing the sentiments to depict the inflation and declination faced by tourism department. These reviews are collected according to the financial years of Tamilnadu tourism and they are compared with the financial growth. This paper is organized as follows: The section II presents the literature survey of this study, Section III explains the research methodology and the section IV focuses on the findings and the final section discusses the conclusion with future research of this study.

II.   RELATED WORK

This section briefs about the related work and literature study regarding sentiment analysis. Karan Chawlaet. al [2] discourses that connectives, conjunctions, negations and conditionals do alter the sentiment of piece of text. The

V. Senthil, Assistant Professor, Department of MBA, Thiagarajar School of Management, Madurai. E-mail:veersenthil@gmail.com

contextual polarity of the sentence depends upon the conjunctions because the adjective describing the noun feeds its effect through the connectives, conjunctions and negations.

Ritesh Agarwal et.al [3] developed an approach to use linguistic resources at phrase level for the analysis at sentence level. They incorporated their approach with support vector machines to conclude that linguistic analysis plays a significant role in sentiment determination.

Marcela Charfuelan et.al [4] correlated between sentiment scores and the acoustic features that are present in the narrative text, such as emotions present in the speech of different characters of a story. Xiaowen Ding et.al [5] formulated linguistic rules and aggregate functions to find the sentiments of the sentences. This analysis uses a semi supervised approach to summarize the product features based on the synonyms of the phrases used by the group of people. Guang Qiu et.al [6] proposed a novel approach to establish relationships between the sentiment words that are collected based on the domain and the features that are modified by those sentiment words. Jens Kr. Steen Jacobsen et.al [9] study reveals a dominance of visual content, along with the relevance of altruistic and community-related motivations and motivational differences between types of content creators. Sharing practices through social media appear as valuable articulations of sociability and emotional support, while having lesser relevance as information sources for holiday decision-making.

A. Tourist behavior

Dorina Maria Buda et.al [7] proposed the conceptual implications when attention turns to emotions and affects, and then describes how tourists and tour guides feel in areas of ongoing conflict. Considering emotions, affects and feelings opens up new research avenues for tourism studies. Benxiang Zeng et.al [8] suggested that research on social media in tourism is still in its infancy. It is critical to encourage comprehensive investigation into the influence and impact of social media (as part of tourism management/marketing strategy) on all aspects of the tourism industry including local communities, and to demonstrate the economic contribution of social media to the industry. A destination image viewed by Zhou et al [10] as a mental projection of tourism imaginary. The social networking sites provide new ways to market destinations and offer diversified platforms to construct destination images.

B. Tourism Marketing

Marketing in Tourism context is defined by Gronroos et al, [11] as ''Marketing is a customer focus that permeates organizational functions and processes and is geared towards making promises through value proposition, enabling the fulfillment of individual expectations created by such promises and fulfilling such expectations through support to customers' value-generating processes, thereby supporting value creation in the firm's as well as its customers' and other stakeholders' processes''. The paradigm of Integrated Marketing Communications (IMC) in digital environment is analyzed by Beba Rakic et al, [12] through five aspects such as integration in terms of traditional media and digital media,

integration in terms of communication methods (Internet communications, mobile communications, communications via digital radio and TV), integration in terms of time of communications and possibilities for interaction, ("monologue" or "dialogue") towards the target audience and dynamic communications in real time, integration in terms of actors (consumers and organizations) and integration in terms of content creation. An overview of medical tourism by John Connell [13] is given of the short history and rapid rise of medical tourism, its documentation, current knowledge and analysis of the industry. John insists that more analysis is needed of the rationale for travel, the behavior of medical tourists, the economic and social impact of medical tourism, the role of intermediaries, the place of medical tourism within tourism (linkages with hotels, airlines, travel agents), ethical concerns and global health restructuring. Louie A. Divinagraciaa et al, [14] stated that the five most important digital communication sources of information about nature-based tourism were search engines; advertisements on the internet; and tourism specific websites such as travel and vacation websites, airline websites, and government tourism websites.

Selling rooms online by Alessandro et al, [15] outlines professional contributions and managerial implications for both the hospitality and Online Travel Agent industries. The author also investigates the relationship between the importance of social media and booking.com using logit class of models. Paris et al, [16] has concentrated on integration of social media into a comprehensive marketing strategy for each destination is the probable best strategy. A conceptual framework by Ruth N. Bolton et al, [17] provides for considering the antecedents and consequences of Generation Y's social media usage. It identifies unanswered questions about Generation Y's use of social media, as well as practical insights for managers. Adventure tourism by Fabian Schlegelmilch et al, [18] continues to be a fast growing and popular form of tourism. Marketing the adventure, however, can present a challenge to small businesses with a limited budget. Viral marketing may present a cost-effective opportunity to reach the youth traveler segment.

C. Electronic Word of Mouth

Digital social media offer low costs per message, unlimited geographical reach and instant access and it is a marketing opportunity to engage the audience in a two-way interaction. Monitoring the social media conversation can help the marketers to expand their outreach and get their message to larger audiences. Kyung-Hyan Yoo et al. [19] observed that the reason for including tourism websites for social media analysis is that access links to social media often are located on the website homepage, targeting tourists or residents.

III. METHODOLOGY

This section describes about the proposed model for efficient sentiment analysis. Initially, the user generated data for tourism places highlighted by Tamilnadu tourism are extracted using corresponding API's such as facebookAPI, twitterAPI respectively. These data are sectionized based on the places i.e., grouping of reviews based on places or tourism

spots. The preprocessing phase of obtained data based on the reviews some features are generated by sentiment detection and sentiment evaluation. For example, if a tweet states "The beach is very clean and has lot of street food outlets plus some rides for children. One can walk and leisured pace and enjoy the scene in Chennai" feature about that particular spot like cleanliness, food availability and entertainment can be evaluated for that spot. Evaluation phase includes the polarity analysis i.e. whether the expressed sentiment by the user is positive or negative. In above example the words "very clean, lot of food, rides, and walk" provides the sentiment of particular user to the place. These words or sentiments are that the user is having a positive idea over the place. But single man's view does not provide the exact picture of the place. So heterogeneous reviews are considered and scores are given to each feature. Now a set of user reviews about a particular place and the scores for the features that are commonly cited in the user's reviews. Figure 1 shows the sentiment analysis model for online reviews and the following are the steps involved in this model,

- Data Pre-processing: In data mining the data should be clean i.e. irrelevant information must be removed in order to provide a focus over the useful data. In this paper, reviews and comments which has been extract from the web are pooled by two factors: (i) place and (ii) year. Unrelated comments and reviews are filtered off.

- Feature Extraction: After filtering off the unrelated content from the data, based on the reviews/opinions of the user the most discussed features are identified and those reviews are grouped according to the similar features. It will be used for the feature based sentiment analysis.

- Part Of Speech tagging: In this step, the sentences in the data set collection are tokenized using the POS tagger of Stanford. During this process, a part of speech such as noun, verb, adverb, adjective, conjunctions, negations and the like are assigned to every word in the sentences. It has been made sure that the conjunctions or negatives present in the sentences are tagged correctly using General Inquirer's word dictionary.

- Conjunction Analysis: A conjunction is a word that is used to link words, phrases, or clauses. It is often used to express the relationship between two clauses or phrases in a sentence. If conjuncts are present in the reviews then those conjunctions are analyzed to find the sentiment of the review.

- Senti-Score: Sentiments are detected for each word using General Inquirer as positive, negative, strong, weak, pleasure, pain and feel. Scores are allotted for the features in the review based on the sentiments obtained by matching the sentiment to the General Inquirer output.

- Sentiment Aggregation: Weighted Average - By the previous block we have got score's for each feature of a review. Now the total score for a review need to be calculated. We used Weighted Average method to calculate the overall sentiment of a review. Class

Label Prediction: The output of the weighted average is subjected to the polarity analysis i.e. tested for positivity and negativity. They are classified by the polarity result and brought under respective class labels.

- Multiple Regression: Multiple regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of two or more variables- also called the predictors.

More precisely, multiple regression analysis helps us to predict the value of Y for given values of X_1, X_2, …,X_k. The detection of the best feature can be found by multiple regression. The variable whose value is to be predicted is known as the dependent variable and the ones whose known values are used for prediction are known independent (exploratory) variables. The class labels are considered to be the dependent variable and the senti-score for the individual features is considered to be the independent variable. In general, the multiple regression equation of Y on X_1, X_2, …,X_k is given by:

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots\ldots\ldots + b_kX_k \qquad (1)$$

Here the independent features (X_1, X_2,….X_k) are road, climate, cleanliness, food, accommodation, vehicle facility, etc… The dependent variable denotes the class label of the particular review/tweet/opinion. The multiple regression results in the identification of the best feature in the tourism spot and the bad feature which needs to be improved. This is applied to found for all the reviews of the tourism spots and the overall best feature discussed on Twitter can be predicted.

- Correlation Analysis: Correlation can tell about the relationship between variables. It is used to understand: whether the relationship is positive or negative and the strength of relationship.

IV. FINDINGS

The growth of tourism business is depends on various factors such as Number of tourists visited (domestic and foreign), Foreign Exchange Earnings, Visa on Arrivals, Aggressive advertisement campaigns on the tourist destinations, Rapid growth of the Indian economy and others. In this study the Growth Rate and the Number of tourists visited (domestic and foreign) are factors taken for analysis. This study is based on secondary data and data reliability is based on the data collecting origination. The most commonly used online materials such as annual reports, websites, online reviews, blogs, online forums and information from social media are taken for a detailed data analysis.

To verify the association between social media results all the Indian states and Union Territories (UTs) official websites (36 numbers) are taken. The score for social media presence is assigned as follows, if the social media tool (facebook, Twitter, YouTube and Flickr) is present and working social media score (sms) two is assigned (sms = 2) and if the tool is present but not working (sms=1) and if the tool is not present (sms= 0). The States & UTs of 19 numbers are scored zero, means they are not having social media presence, the remaining 17 states scored nonzero value, i.e. these states have

the social media presence in their official websites. The Telungana is a new state; it is not included in this study. The major observation during this verification process is, less social media awareness among the Union Territories and North Eastern states (except Aurnachal Pradesh whose sms is 4 out of 8). The other observations are Andaman and Nicobar Islands (UT) has Short Message Service (SMS) alert facility, Assam has Google Translator facility and West Bengal has WB lens a video gallery facility in their official websites. Figure 2 shows the pie graph of facebook interactions and Figure 3 shows the twitter interactions created by the Firm Generated Content (FGC) i.e. by the state DMOs. Kerala State is topper in this list, it has an excellent social media marketing strategies, and it was the first one joined (21/1/2010) in Social Media as a marketing channel. Kerala DMO has the maximum number of interactions in social media (facebook, Twitter and Youtube) except "talking about this" in Facebook where Gujarat is topper. Also from the observed results, Kerala DMO is updating the content regularly and frequently in social media. Kerala had recently launched a host of path-breaking mobile-based services and IT-enabled projects to further enhance travel experience of visitors. The new set of products include mobile website, WAP guide, applications for Android and IOS and Bluetooth Kiosks. Kerala received the national award (2012) for the most innovative use of information technology from the Ministry of Tourism, India.

For micro and in depth data analysis, we have proposed the model to analyze the tweets and opinions on ten places in Tamilnadu such as Chennai, Madurai, Coimbatore, Yercaud, Kodaikanal, Ooty, Kaniyakumari, Trichy, Rameshwaram, Theni. We have used General Inquirer to support the POS tagging process. During the training phase, we have identified the rules for sentiment orientation considering the presence of conjuncts, to improve the accuracy of decision on tweets and opinions.

Example 1: "This place is beautiful but the road is bad" In this sentence the first clause is positive however the clause following the conjunction has a negative effect. So the overall sentiment expressed in this sentence is negative. It has one positive word "beautiful" (+1), one negative word "bad" (-1). So the overall senti-score for the feature "road" is -1 (negative). We collected the reviews on various features of the tourism spots and they are used in two different approaches.

Feature Extraction process results are listed in Table 1, after the Feature Extraction process, the POS Tagging and General Inquirer in the Sentiment detection is done and the score are calculated based on the sentiment of the features. And for each feature the scores are given, finally the overall senti-score of the review is calculated. The class label is predicted based on the senti-score and the results are listed in Table 2.

Input (Tweets/reviews/Comments)
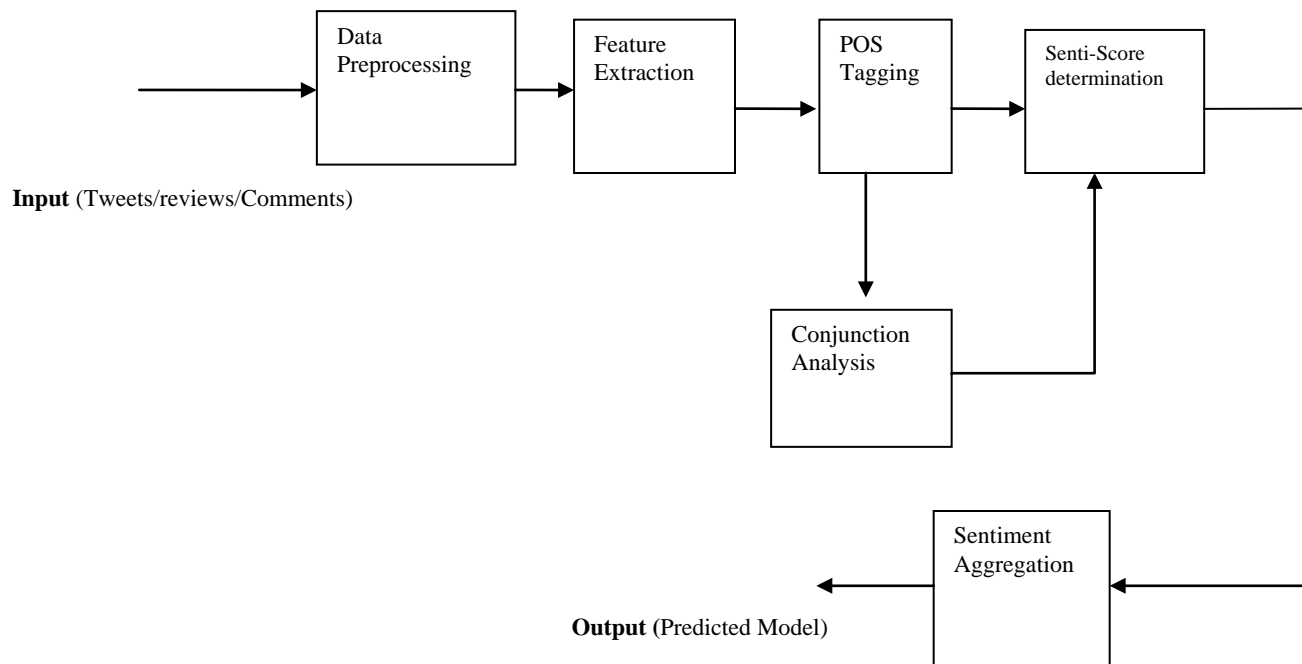
Output (Predicted Model)

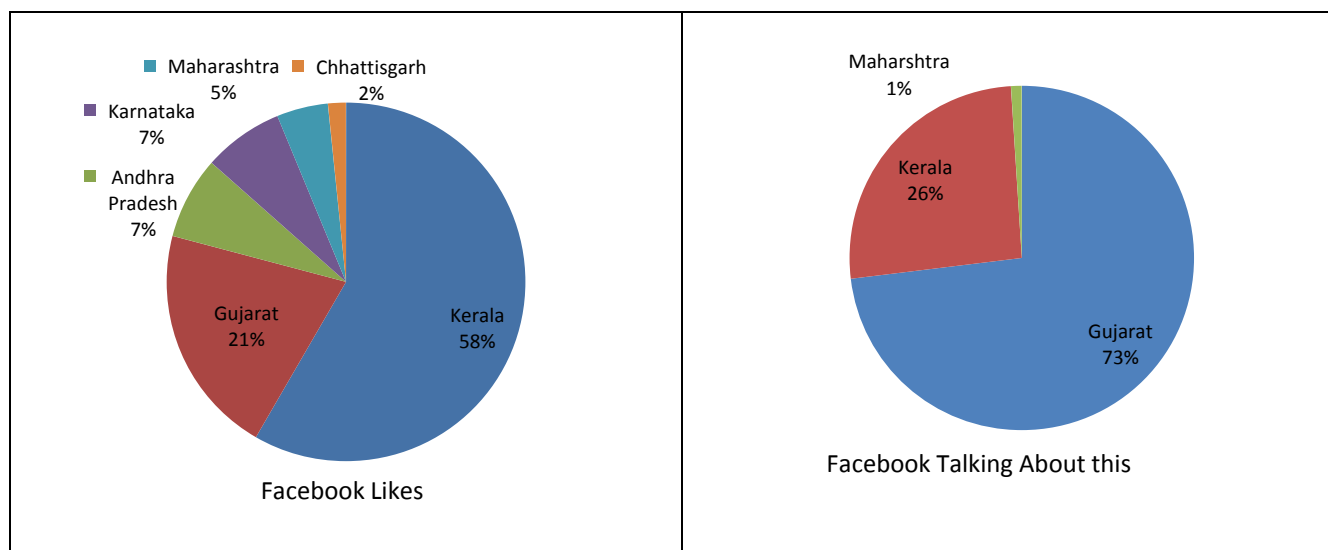Fig. 1: Sentiment Analysis Model for Online Reviews

Fig. 2: Facebook Interactions by the State DMOs
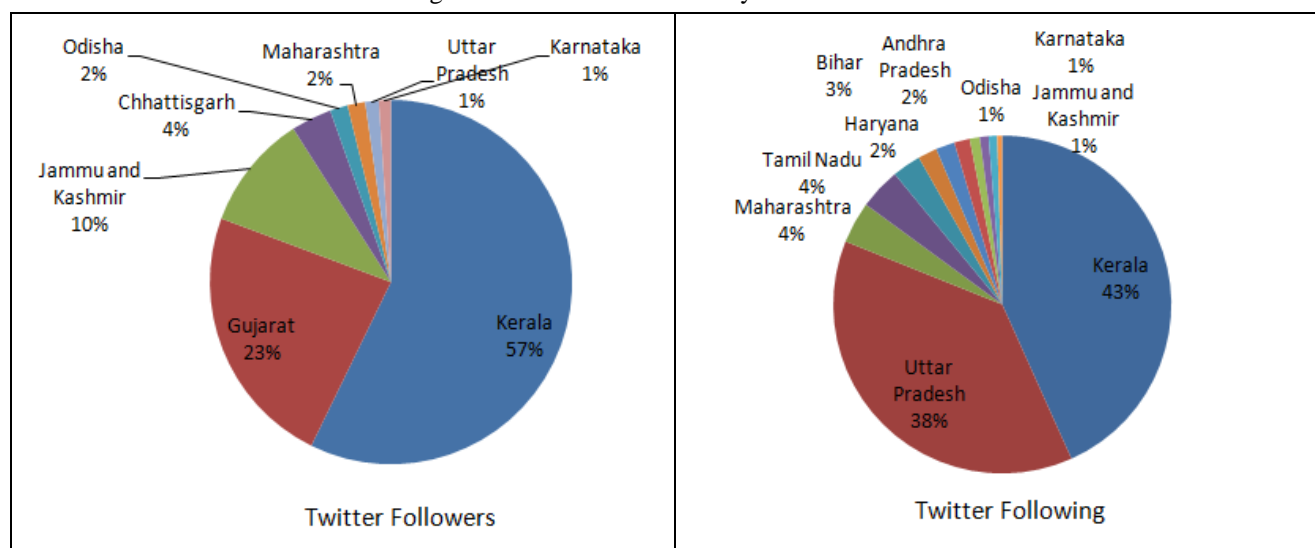
Fig. 3: Twitter Interactions by the State DMOs

Table I: Feature Extraction

Sno	Reviews	Features Extracted				
1	The beach is very clean and has lot of street food outlets plus some rides for children.	Clean	Food	Ride	NA	NA
2	Cool breezes excite us to stay. There is a besantnagar beach too which is very unclean and poor crowd compared to Marina.	Unclean	NA	NA	crowd	NA
3	Maintaining Cleanliness is needed and patrolling of Security to save the public from unsocial element Multiple fooding outlet.	Clean	Food	NA	Crowd	Security

Table II: Sentiment Detection

Review	Cleanliness	Food Availability	Crowd management	Security	Ride	Senti-score	Class label
1	3	2	0	0	2	7	positive
2	-3	2	-2	0	0	-3	negative
3	2	3	0	4	0	9	positive

V.   CONCLUSION AND FUTURE ENHANCEMENT

In this paper, the growth of Tamilnadu tourism business is influenced by social media interactions are analyzed. The interactions based on the Firm Generated Content such as Facebook (Likes, Talking about This) and Twitter (Tweets, Followers, Following) and Tourist Generated Content are taken for analyze the growth of the Tourism business. The research results revealed that Indian DMOs are slower in adapting the social media in their business. Since the social media is low or zero cost medium of marketing channel and also popular among the youths, the Indian DMOs should use this opportunity to promote their business. Since this study is an early effort, the separation of growth rate by social media

from the regular marketing channel is not analyzed, further study is required over the DMOs budget and Return On Investments. The sentiment analysis by TGC will be helpful to DMOs and for Tourists for two way interactions. Finally, this study offers valuable insights of social media into researchers. Also social media interactions are helpful to tourists and finally provide better business to state DMOs. The user generated data provides a higher degree of accuracy about what exactly the user feels about a particular tourism place. Feature based analysis provided what features in a particular place has positive and negative view by the users. The feature which has strong negative sentiment is a suggestion to the tourism department so that they can improve their Infrastructure facility and this leads to revenue improvement. The future research work will include Questionnaire and Interview methods to analyze the growth of the Indian Tourism business.

REFERENCES

[1] Anandhi Bharadwaj, Omar A. EI Sawy, Paul A. Pavlou, N. Venkataraman, "Digital Business Strategy : Toward a next generation of insights", MIS Quarterly Vol. 37, No. 2, pp 471-482, June 2013.

[2] Karan chawla, Ankita Ramteke, Pushpak Bhattacharyya, "IITB-Sentiment-Analysts: Participation in sentiment analysis in Twitter" , Joint Conference on Lexical and Computational Semantics, 2013

[3] Ritesh Agarwal, "I Know what you feel": Analyzing the role of conjunctions in Automatic Sentiment Analysis IIT Kanpur, GoTAL LNAI 5221, pp 28-30, 2008, Springer Velag Berling Heidelberg 2008.

[4] Marcela Charfuelan, Marc Schroder, Correlation analysis of sentiment analysis scores and acoustic features in audiobook narratives, in 4thInternational Workshop on Corpora for Research on Emotion Sentiment & Social Signals, Istanbul, Turkey, 2012.

[5] Xiaowen Ding and Bing Liu "The Utility of Linguistic Rules in Opinion Mining",SIGIR '07 Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Pages 811-812

[6] Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen, "Expanding Domain sentiment lexicon through double propagation", Proceeding IJCAI'09 Proceedings of the 21st international joint conference on Artifical intelligence Pages 1199-1204.

[7] Dorina Maria Buda, Anne-Maria d'Hauteserre, Lynda Johnston, "Feeling and tourism studies", Annals of Tourism Research, Volume 46, May 2014.

[8] Benxiang Zeng, Rolf Gerritsen, "What do we know about social media in tourism? A review", Tourism Management Perspectives, Volume 10, April 2014.

[9] Jens Kr. Steen Jacobsen, Ana Maria Munar, "Motivations for sharing tourism experiences through social media", Tourism Management, Volume 43, August 2014.

[10] Zhou,L., " Online rural destination images: Tourism and rurality", Journal of Destination Marketing & Management, Elsevier (2014).

[11] Grönroos, C. "On defining marketing: finding a new roadmap for marketing", Journal of Marketing Theory (2006) 6, 395–417.

[12] Beba Rakic, Mira Rakic, " Integrated Marketing Communications Paradigm In Digital Environment : The Five Pillars Of Integration", Scientific review paper, Vol. II, No I, (2014) 187-204.

[13] John Connell,"Contemporary medical tourism: Conceptualisation, culture and commodification", Journal of Tourism Management 34 , 1-13, 2013.

[14] Louie A. Divinagraciaa*, Maria Rochelle G. Divinagraciaa, David G. Divinagracia, "Digital Media-Induced Tourism: The Case of Nature-based Tourism (NBT) at East Java, Indonesia", Procedia - Social and Behavioral Sciences 57 (2012), 85 – 94.

[15] Alessandro Inversini, Lorenzo Masiero, "Selling rooms online: the use of social media and online travel agents", International Journal of Contemporary Hospitality Management, Vol. 26 No. 2, (2014), pp. 272-292.

[16] Paris, C. M. (2012). Flashpackers: An emerging sub-culture? Annals of Tourism Research, 39(2), 1094–1115.

[17] Ruth N. Bolton, A. Parasuraman et. al. "Understanding Generation Y and their use of social media: a review and research agenda", Journal of Service Management, Vol. 24 No. 3, 2013, pp. 245-267.

[18] Fabian Schlegelmilch and Claudia Ollenburg, "Marketing the adventure: utilizing the aspects of risk/fear/thrill to target the youth traveler segment", Tourism Review, vol. 68, no. 3 2013, pp. 44-54.

[19] Kyung-Hyan Yooa, Jangyul Robert Kim, "How U.S. state tourism offices use online newsrooms and social media in media relations", Elsevier Journal of Public Relations Review 39 (2013), 534– 541.

[20] http://nlp.stanford.edu/software/tagger.shtml

[21] http://www.wjh.harvard.edu/~inquirer

Dr.V. Senthil is working as Assistant Professor in the Department of Management Studies, Thiagarajar School of Management, Madurai, India. He has 16 years of Teaching experience, 7 years of research and 3 years of Industry experience. He has published more than 22 research articles in the International & National Conferences and Journals. He has also presented 2 research articles in UAE and Malaysia. He has chaired many technical sessions and delivered invited talks in National and International Conferences. He has awarded and completed two Summer Faculty Research Fellowships one at IIT, Delhi (2014) and other at ISI, Kolkatta (2009). He is a member of Advanced Computing Society and Computer Society of India. His research interests are: Social Media Analytics, Cloud Computing, Big Data, Data Mining and Digital Video Watermarking.