International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12431-12439

Opinion Mining and Sentiment Analysis of Travel Websites through Twitter

Piyush Bhardwaj[1], Dr. Suruchi Gautam[2] and Dr. Payal Pahwa[3]

[1]Research Scholar, Department of Computer Science & Engineering, Uttarakhand Technical University, Dehradun, Dehradun, Uttarakhand, India.
[2]Associate Professor, Department of Computer Science, Rajdhani College, University of Delhi, New Delhi, India.
[3]Professor, Department of Computer Science & Engineering, Bhagwan Pparshuram Institute of Technology, Guru Gobind Singh Indraprastha University, New Delhi, India.
[1]Orcid Id: 0000-0002-2545-0505

Abstract

Opinion mining and sentiment analysis are the techniques that analyse the opinions and sentiments of any sentence. This analysis can be used for decision-making and strategic planning. In this paper, we have used these techniques on one of the biggest service industry in the world: travel industry. The motivation of this work comes from the recent upsurge in travelling. This work focuses on eleven countries whose citizens travel the most. These people use various online platforms to plan their travels. There is a massive competition among these major online companies to become a global leader in travel industry. Our research identifies the most trusted travel company in the world. It also calculates the most talked about company among all companies in various countries. The data source used is Twitter and research is conducted on the basis of our approach, which is based on lexical analysis and Naïve Bayes algorithm. The analysis is done in statistical package R. According to our results, Airbnb is the most preferred as well as most trusted travel website among all the market leaders.

Keywords: Lexical analysis, Naïve Bayes algorithm, travel websites, sentiments.

INTRODUCTION

Travelling is driven by two sides of the same coin. It is either for work or a break away from work. Travelling is supposed to bring satisfaction to the traveller in accordance with the purpose of his travel. Both kind of travellers look for comfort and ease in the process of travelling so that they can focus on the purpose of their travel. A traveller moving for work reasons is usually assisted with bookings and itinerary. Such a traveller has the least choice or interest in the destination or the experience. His focus is productivity and not pleasure. He is the least emotional in his travel experience because even while travelling work occupies his thoughts.

The next category of travellers is more engaged and emotional about the travel experience. People who save money to travel, who wish to make their journey a memorable and pleasurable experience are also constantly checked by the cost for this experience. They want the maximum out of the least expenditure. This pushes them to use social media.

The social media then becomes a platform for people to share their experiences and learning with regard to itineraries, expenditure, things to do, things to see, experiences with specific places and people and finally the benefit of travel websites promising to assist travel experience. While social media like facebook, instagram offer unlimited space and scope of sharing experiences, a platform like twitter restricts long narratives. Twitter allows the users to express themselves in 140 characters for each tweet. This format caters to the reducing attention span of the social media consumers. People spend a lot of time researching on travel websites like tripadvisor, expedia, Airbnb etc. However, when it comes to giving feedback on the utility of these websites, twitter comes in handy for people to share their sentiments and opinions. Twitter becomes a platform for people to quickly post and receive leads about where to find the most helpful assistance with regard to travelling issues. Because of the restricted nature of this platform, it is helpful to both the consumers and the service providers like these travel websites. It helps the websites to improve on their system and also the travellers to go to the most helpful website without devoting time in comparing all of them. For the same reasons, twitter helps a researcher to make a successful analysis of sentiments and opinions expressed in tweets. This analysis can be used for decision-making and strategic planning.

In this paper, we have performed opinion mining and sentiment analysis on one of the biggest service industry in the world: Travel Industry. The work focuses on eleven countries whose citizens travel the most. These people use various online platforms to plan their travels. Our research identifies the most trusted travel company in the world. It also calculates the most talked about company among all companies in various countries. Research is conducted on the basis of our approach, which is based on lexical analysis and Naïve Bayes algorithm. This research is relevant to the travellers because it helps them understand travel websites on the basis of quotients like popularity, positivity and emotions. Therefore, it will help the researchers to save time and energy, make their searches efficient and focused The research will also help travellers choose a website with the best performance in a specific region. The research helps the travel industry to self analyze their processes and products. The travel industry benefits from this research by learning their weaknesses in terms of location, positivity and emotions.

The paper is organized as follows: Section 2 provides the related work in sentiment analysis. Section 3 explains the methodology used for sentiment analysis through our system. Section 4 analyses the results. Section 5 is the concluding section. Finally future scope is given in Section 6.

RELATED WORK

The author in [1] proposes an aspect-level sentiment analysis method based on ontologies in the diabetes domain. For calculating efficiency of their method, they used a Twitter corpus. They concluded that N-gram around method was the best. The author [2] develops a Machine leaning algorithm that exploits all hash tags and emoticons inside a tweet. The system was proved efficient, robust and scalable.

In paper [3], the authors identified a simple and workable approach for Arabic sentiment analysis on Twitter. They used three techniques, Naive Bayes, Support Vector Machine and Decision Tree. Their results conclude that Decision Tree is better then the rest of two algorithms. Paper[4] researched on Roman-Urdu opinion mining using three known algorithms, Naive Bayes, Decision Tree and KNN using WEKA tool. Opinions were extracted from a blog. Their results concludes that Naive Bayes is better than Decision Tree and KNN in terms of more accuracy, precision, recall and F-measure.

Paper [5] proposes sentiment analysis for hotel. Their results proved support vector machines as the most accurate algorithm. In [6], the author discusses trend and sentiment analysis of IoT from multiple Twitter data sources and validated these trends with Google Trends. The author concludes that people were positive about IOT.

In [7], the author studies tweets of advertisements of electronic cigarette. They concluded decrease in positivity of Organic tweets as compared with automated tweets. In [8], the author proposes a prototype system for short text understanding. The author uses semantic knowledge provided by a knowledge base and harvested from a web corpus. Their knowledge-intensive approaches unsettle the traditional methods for performing tasks such as part-of-speech tagging, concept labeling and text segmentation.

[9] formalized Social Observatory for observing and measuring social indicators. The authors analysed 54,665 posts and 231,147 comments. The authors were able to conclude how users interacted, with whom and at what volume. [10] provided an emoji sentiment lexicon, called the Emoji Sentiment Ranking and draws a sentiment map of the 751 most frequently used emojis. The author also proposes Emoji Sentiment Ranking. [11] Determines collective sentiment for climate change news, events, and natural disasters. They concluded that natural disasters, climate bills, and oil- drilling decrease in happiness while climate rallies, a book release, and a green ideas increase in happiness. In the paper [12], the author identifies and discusses the opportunities that are growing for cross-disciplinary work, which will enhance individual advances. The author presents s comparative analyses of the sentiment detection methods and sentiment-related phenomena used in both communities and gives an overview of the aims of socio-affective human-agent strategies.The [13] studied people's publishing and tweeting frequency, use of hash tags, language, and emotions Twitter. Their analysis shows that astrophysicists address the different groups but that they do not talk to each. Survey [14] gives a detailed analysis of sentiment analysis techniques and the related fields with brief details. [15] Studied tweets from Kidney Week 2011 to increase public awareness of kidney disease. Study [16] monitor usage of terms ''H1N1'' versus ''swine flu'' over time, analysis of ''tweets''; and validate Twitter as a real-time content, sentiment, and public attention trend-tracking tool. Tweets can be used for real-time content analysis and knowledge translation research, allowing health authorities to respond to public concerns.

PROPOSED APPROACH

After Our approach for analyzing sentiments on a given data source uses a combination of lexical analysis and Naïve Bayes algorithm [17]. A lexical based approach is an unsuperwised approach for sentiment analysis. This approach relies on external sources to calculate the polarity of a sentence. Naive Bayes algorithm is used to analyze multi feature data. Naïve bayes is a probabilitic classifier that returns a class with maximum postirier probability of the document. Naive Bayes algorithm is used to analyze multi feature data. We have defined our Bayes classifier for calculating the probability of a positive tweet on the basis of Bayes rule A confusion matrix test of the approach gives an accuracy of more than 83 percent. The analysis is done in statistical package R. Three parameters are taken for checking the polarity of a stream. POS, a positive parameter that tells whether the stream tends to have a positive point of view, NEG, a negative parameter that tells whether the stream tends to have a negative point of view and NEUTRAL, a parameter that tells whether the stream tends to have a neutral point of view. Positive score is calculated by adding one to value of Pos everytime a positive word is encountered in the tweet. Negative score is calculated by adding one to Neg whenever a negative word is encountered in the tweet. Neutral score is calculated by subtracting Pos and Neg. Emotions are categorized in six categories named anger, disgust, fear, joy, sadness and surprise. Values of all parameters are calculated on the basis of the formulas mentioned below. We have taken 10 major companies for our consideration. They are Airbnb, Expedia, Hotwire, Kayak, Lonely Planet, Orbitz, Skyscanner, Travelocity, Tripadvisor, Viator. To move further in our research, we took 11 countries from the world whose people travel the most [18]. They are Finland, USA, Sweden, Denmark, Norway, Hong Kong, New Zealand, Canada, Australia, France, and India. The approach is applied on every country for every company respectively. A time stamp of 50 days is taken. The tweets are extracted from 1st Feb 2017 to 20th Mar 2017 with a cap of 500 tweets for a particular run. The equations used for the analysis are as follows:

$$\mathrm{Pos}(k) = \sum_{j=1}^{L} pw_j \tag{1}$$

$$\mathrm{Neg}(k) = \sum_{j=1}^{L} nw_j \tag{2}$$

$$\mathrm{Neu}(k) = \mathrm{Pos}(k) - \mathrm{Neg}(k) = \sum_{j=1}^{L} pw_j - \sum_{j=1}^{L} nw_j \tag{3}$$

$$P(\,\mathrm{Pos} \mid \mathrm{Tweet}) = \frac{P(\mathrm{Pos})P(\mathrm{Tweet} \mid \mathrm{Pos})}{P(\mathrm{Tweet})} \tag{4}$$

$$P(\mathrm{Tweet} \mid \mathrm{Pos}) = P(\,w_1 \mid \mathrm{Pos}) * P(w_2 \mid \mathrm{Pos}) * \ldots * P(\,w_n \mid \mathrm{Pos}) \tag{5}$$

$$P(\,\mathrm{Neg} \mid \mathrm{Tweet}) = \frac{P(\mathrm{Neg})P(\mathrm{Tweet} \mid \mathrm{Neg})}{P(\mathrm{Tweet})} \tag{6}$$

$$P(\mathrm{Tweet} \mid \mathrm{Neg}) = P(\,w_1 \mid \mathrm{Neg}) * P(w_2 \mid \mathrm{Neg}) * \ldots * P(\,w_n \mid \mathrm{Neg}) \tag{7}$$

$$P(\mathrm{Sentiment} \mid \mathrm{Tweet}) = \frac{P(\mathrm{Sentiment})P(\mathrm{Tweet} \mid \mathrm{Sentiment})}{P(\mathrm{tweet})} \tag{8}$$

$$P(\mathrm{Tweet} \mid \mathrm{Sentiment}) = P(w1 \mid \mathrm{Sentiment}) + P(w2 \mid \mathrm{Sentiment}) + \cdots + P(w_n \mid \mathrm{Sentiment}) \tag{9}$$

Equation 1,2 and 3 calculates the polarity of the tweets. Pos(k) yields positivity of a tweet. Neg(k) yields negativity of tweets. Neu(k) yields value of neutral tweets. Positive score is calculated by adding one to value of Pos everytime a positive word is encountered in the tweet. Negative score is calculated by adding one to Neg whenever a negative word is encountered in the tweet. Neutral score is calculated by subtracting Pos and Neg. In the above formula's, L is the length of words in the tweet, k is set of countries. Equation 4,5,6,7 calculates positive and negative tweets through Naïve Bayes classifier. P(Pos | Tweet) is the probability of positive tweets. The two probabilities P(Pos) and P(Tweet) are independent of all other probabilities. In order to find the probability of P(Tweet | Pos) we use equation 5. P(Neg | Tweet) is the probability of negative tweets. The two probabilities P(Neg) and P(Tweet) are independent of all other probabilities. In order to find the probability of P(Tweet | Neg) we use equation 7. Equation 8,9 calculates the sentiments of the tweet. As sentiments are classified in six categories, anger, disgust, fear, joy, sadness and surprise, equation 8 and 9 are used for all 6 categories respectively.

RESULT ANALYSIS

Our system is executed for 11 countries mentioned in table 1. The Longitude and Latitude of each corresponding country is also mentioned in the table. These are system-generated values retrieved from our approach. The approach is executed for 10 major travel companies in the world. The time duration taken for execution is from 1st Feb 2017 to 20th Mar 2017. Table 2 provides a country wise list of tweets extracted for all the companies in the given duration of time. Table 3 gives polarity of tweets for all the companies. Table 4 contains average Emotions of all companies.

Table 1: Latitude and Longitude of the countries taken into consideration.

Country\location	Longitude	Latitude
Finland	25.74815	61.92411
USA	-95.71289	37.09024
Sweden	18.6435	60.12816
Denmark	9.501785	56.26392
Norway	8.468946	60.47202
Hong Kong	114.1095	22.39643
New Zealand	174.886	-40.90056
Canada	-106.3468	56.13037
Australia	133.7751	-25.2744
France	2.213749	46.22764
India	78.96288	20.59368

The above table presents a Longitude and Latitude of all the countries where our research have been carried out. The tweets are extracted based on these given locations for all the countries.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12431-12439
© Research India Publications. http://www.ripublication.com

Table 2: A country wise list of tweets extracted for all the companies in the given duration of time.

Site\Country	Finland	USA	Sweden	Denmark	Norway	Hong Kong	New Zealand	Canada	Australia	France	India
Skyscanner	1	100	9	166	159	13	2	6	11	173	15
Tripadvisor	19	500	51	500	500	56	6	25	27	500	192
Expedia	13	500	110	500	500	29	24	94	44	500	93
Travelocity	0	494	2	5	4	4	0	4	10	4	6
Kayak	12	500	27	302	283	46	9	91	53	263	79
Orbitz	0	203	0	22	22	6	0	2	3	22	4
Hotwire	0	97	1	51	47	6	0	2	2	24	8
AirBnB	85	500	254	500	500	225	38	152	14	500	500
Lonely Planet	13	313	37	245	240	16	6	22	14	244	22
Viator	1	260	1	63	8	2	0	0	1	60	14

The above table shows a country wise list of total tweets extracted from 11 countries and 10 websites. Airbnb leads the count in all the countries having maximum number of tweets followed by Expedia and tripadvisor. The least tweets extracted are for orbitz and hotwire.

Table 3: Polarity of tweets for all the companies

Companies/Polarity	Positive	Negative	Neutral
Airbnb	2475	613	312
Expedia	1528	628	251
Hotwire	140	66	31
Kayak	1169	296	200
Lonely Planet	227	569	373
Orbitz	185	56	43
Skyscanner	453	112	89
Travelocity	366	129	32
Tripadvisor	1678	345	353
Viator	310	96	67

Table 4: Average Emotions of all companies

Companies/Emotions	ANGER	DISGUST	FEAR	JOY	SADNESS	SURPRISE
Airbnb	3.92	3.14	2.16	2.50	2.06	4.66
Expedia	5.16	3.42	2.11	2.08	2.28	3.84
Hotwire	4.64	3.24	2.27	2.20	2.34	5.34
Kayak	3.46	3.12	2.21	2.10	2.25	4.87
Lonely Planet	3.39	3.11	2.11	1.52	1.82	5.86
Orbitz	4.14	3.14	2.23	2.91	2.02	4.41
Skyscanner	4.08	3.13	2.13	1.87	2.25	4.10
Travelocity	3.79	3.15	2.18	2.12	1.90	5.14
Tripadvisor	3.12	3.14	2.52	2.47	1.95	4.21
Viator	2.96	3.11	3.75	2.25	2.03	5.47

The above table shows average emotions of people for all 10 websites. Emotions are classified into 6 categories and an average of all the emotions for all the countries are calculated for each website. The emotions reveal that on an average, expedia faces the most anger and disgust among all the people with an average value of 5.16. Anger and disgust emotions mostly come from negative tweets. An emotion of fear in highest in Viator followed by Tripadvisor. Airbnb records the highest average emotion of Joy among all the websites with a value of 2.50. A surprise emotion has the highest value among all the emotions. Lonely Planet leads this emotion with an average value of 5.86 followed by Viator (5.47), Hotwire(5.34) and Travelocity(5.14)

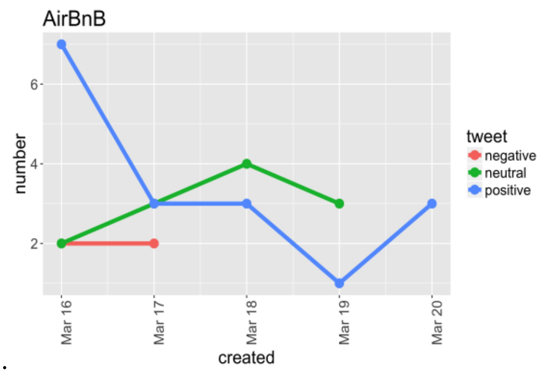


Figure 1: Sample of automated graph generated for Airbnb in New Zealand

The above figure is an automated graph generated from our system to reveal a polarity opinion for Airbnb in New Zealand. The results shows more positive tweets then negative and neutral on various occasions.

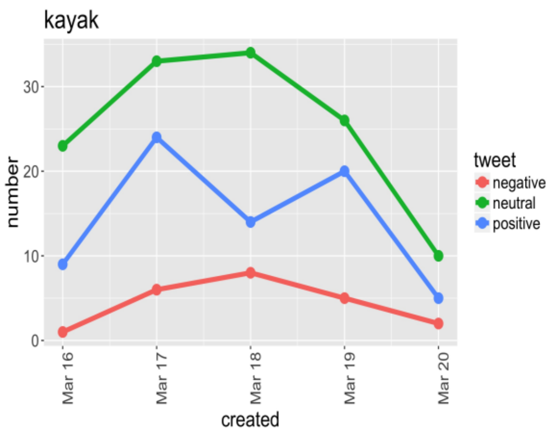


Figure 2: Sample of automated graph generated for Kayak in France.

This automated graphs shows more neutral tweets for kayak in France from 16th March 2017-20th March 2017. Neutral tweets are followed by positive tweets and then by negative tweets.

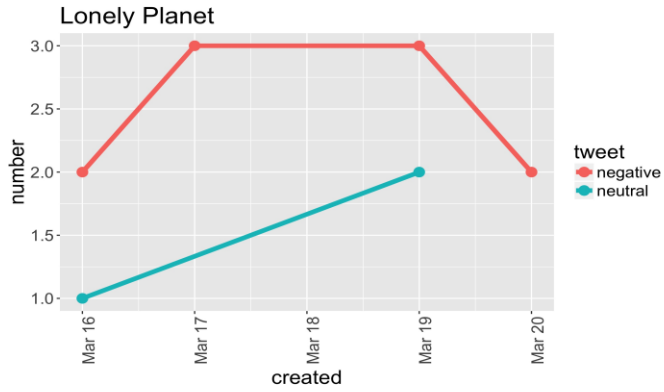


Figure 3: Sample of automated graph generated for Lonely Planet in Canada.

This graph shows that there are no positive tweets for Lonely Planet in Canada between 16th March and 20th March 2017.

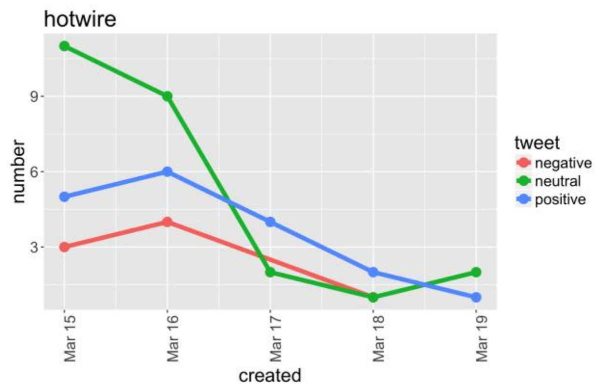


Figure 4: Sample of automated graph generated for Hotwire in Denmark.

The above graph depicts that there is a close proximity between positive and neutral tweets for Hotwire in Denmark in the given time frame. Negative tweets are the least.

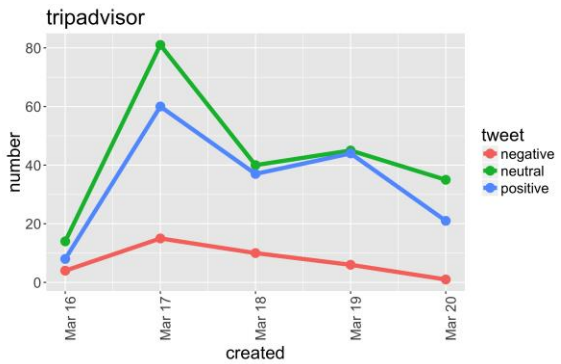


Figure 5: Sample of automated graph generated for Tripadvisor in USA.

In USA, a tripadvisor's polarity shows a similar, number of positive and neutral tweets, for the given time frame. Negative tweets are the least.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12431-12439

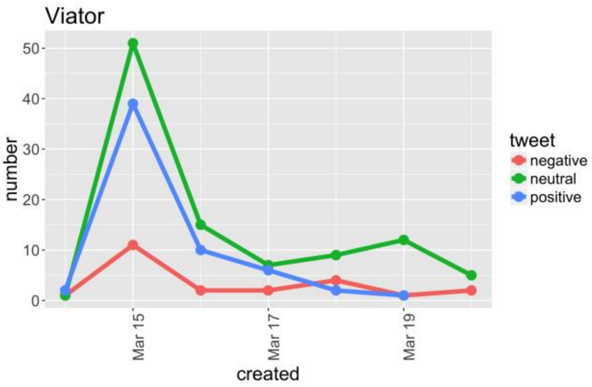


Figure 6: Sample of automated graph generated for Viator in USA.

The above-automated graph depicts the polarity of tweets for Viator in USA for a given time frame. The graph is self-concluding with results showing positive tweets overpowering negative tweets.

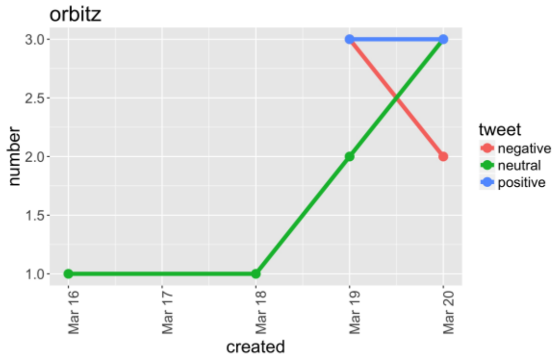


Figure 7: Sample of automated graph generated for Orbitz in Norway.

The above graph presents an opinion for Orbitz in Norway. It shows a maximum amount of positive tweets on 19th March 2017.

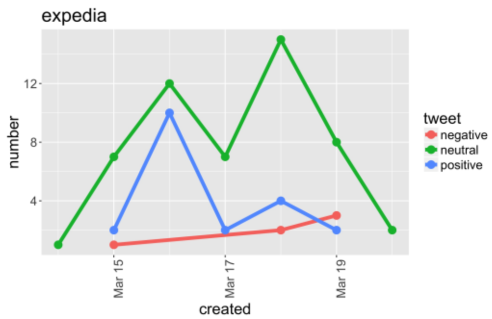


Figure 8: Sample of automated graph generated for Expedia in Swedon.

The above graph concludes the polarity of tweets for expedia in Swedon. A very minimal negative tweets and a good number of positive and neutral tweets makes Expedia a more stronger company to search for.

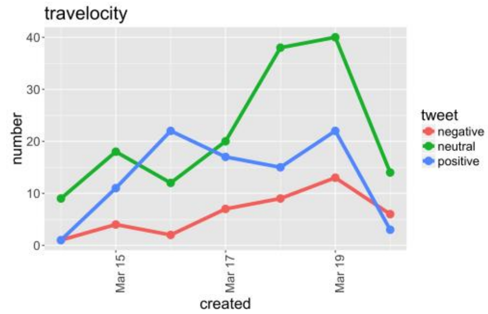


Figure 9: Sample of automated graph generated for Travelocity in USA.

The above graph concludes the polarity of tweets for travelocity in USA. Very few negative tweets and a good number of positive and neutral tweets gives a positive opinion abour Travelocity in USA.

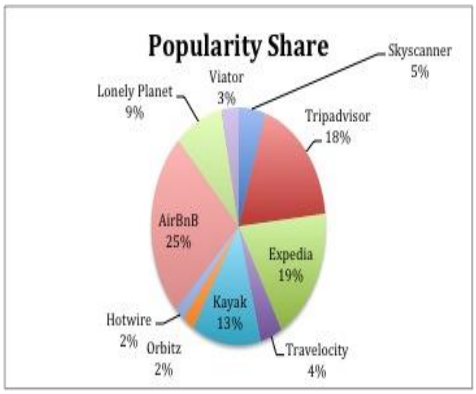


Figure 10: Popularity Share of Travel Companies on Twitter

The above figure is a percentage wise representation of the popularity share of online companies dealing in travel industry. This popularity share has been drawn from our data source: twitter. Airbnb is the most popular company with 25% of the share. Airbnb dominates being 6% ahead of expedia which holds 19% of the popularity share. However, tripadvisor is right behind expedia with a minor 1% gap at 18%. Kayak, at 13%, is the next in holding a considerable share in this representation. Lonely planet(9%), skyscanner(5%), travelocity(4%) don't seem to compete with these big players that put together make 75% of the total popularity share. Viator(3%), hotwire(2%) and orbitz(2%) are the tail enders in this division of popularity. This pie chart gives us a clear picture of the dominant market players. While the competetion seems difficult of the face of it, the gap between Airbnb and Kayak is big enough to hold 4 small companies in between.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12431-12439

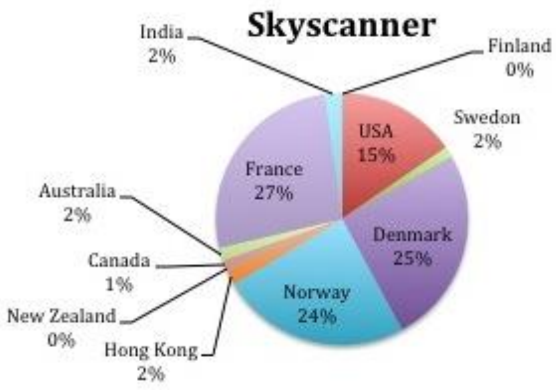


Figure 11: Popularity share of Skyscanner on twitter

The above pie chart depict the popularity share of Skyscanner on twitter of 11 countries whose people travel the most. Skyscanner is most famous in France(27%) followed by Denmark(25%) and Norway(24%). USA is next with a 15% of skyscanner's popularity share on twitter. India, Australia and Hong-Kong trail behind with a mere 2% each popularity share. People in New Zealand and Finland do not find anything to discuss about skyscanner on twitter.

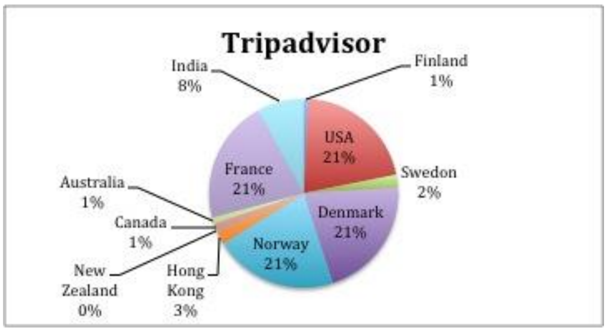


Figure 12: Popularity share of Tripadvisor on twitter

The above pie chart explains the popularity share of Tripadvisor on twitter. It is most famous in France(21%), Denmark(21%),Norway(21%) and USA(21%). India is next with an 8% of tripadvisor's popularity share on twitter followed by Hong-Kong(3%), Sweden(2%), Finland(1%), Canada(1%). People in New Zealand do not find anything to discuss about tripadvisor on twitter.

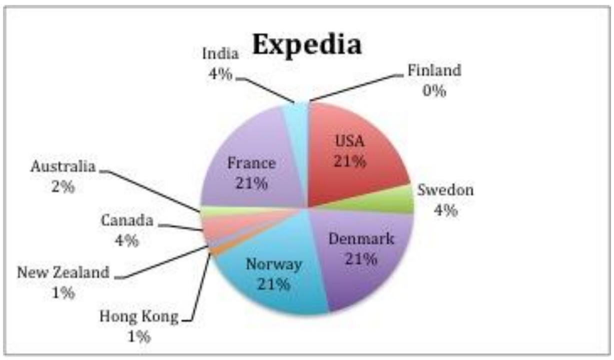


Figure 13: Popularity share of Expedia on twitter

The above chart explains the popularity share of Expedia on twitter. Just like tripadvisor, expedia is most famous in France(21%), Denmark(21%), Norway(21%) and USA(21%). India, Sweden and Canada are next with 4% each of expedia's popularity share on twitter followed by Australia (2%), Hong-Kong (1%) and New-Zealand (1%). People in Finland do not find anything to discuss about expedia on twitter.

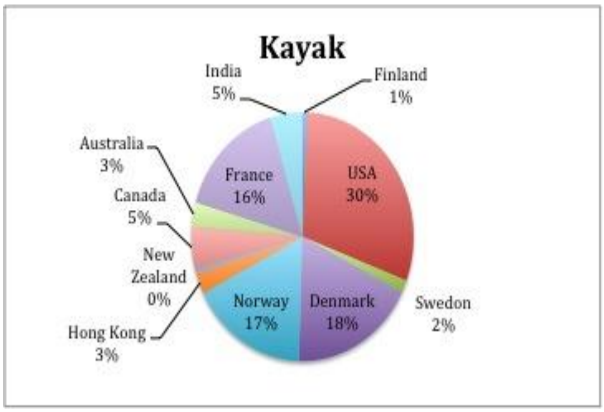


Figure 14: Popularity share of Kayak on twitter

The above chart depicts the popularity share of Kayak on twitter. Kayak is most famous in USA(30%) followed by Denmark(18%), Norway(17%) and France(16%). India and Canada comes next with 5% each popularity share. The tail enders are Australia (3%), Hong-Kong (3%), Sweden (2%), Finland (1%) and New-Zealand (0%).

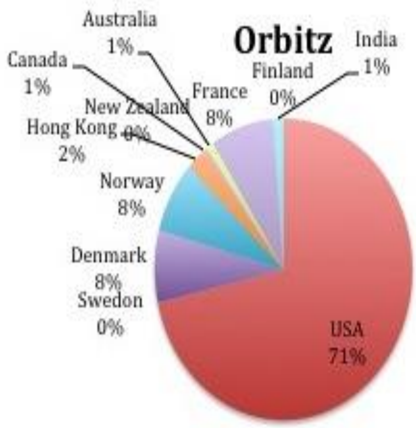


Figure 15: Popularity share of Orbitz on Twitter

The above chart represents the popularity share of Orbitz. It is most famous in USA with an enormous 71% of popularity share. All other countries trail behind with a very low popularity share.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12431-12439

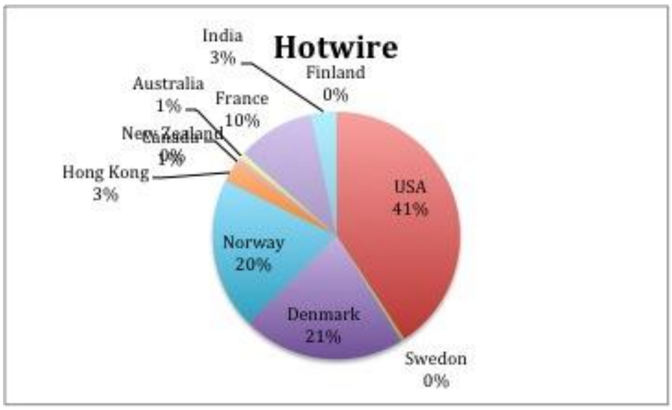


Figure 16: Popularity share of Hotwire on Twitter

The above chart depicts Hotwire's popularity share on Twitter. Just like Orbitz, it is most famous in USA (41%) but has a considerable amount of share in Denmark (21%), Norway(20%) and France(10%). Rest all countries are negligible.

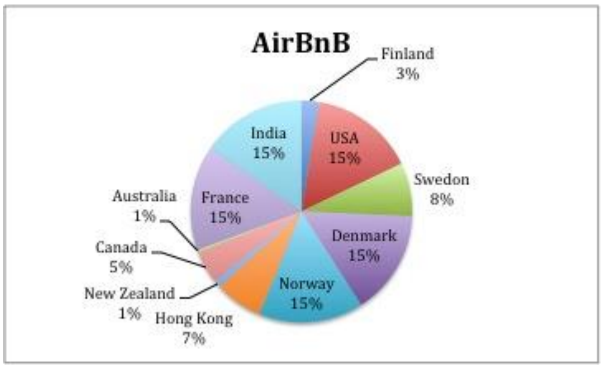


Figure 17: Popularity share of Airbnb on twitter

The chart depicts Airbnb's popularity share on twitter. Airbnb has the most evenly distributed popularity share all across the world. USA, France, Denmark, Norway and India share equal (15%) popularity share for Airbnb. Sweden (8%), Hong-Kong(7%) and Canada(5%) are not far behind.

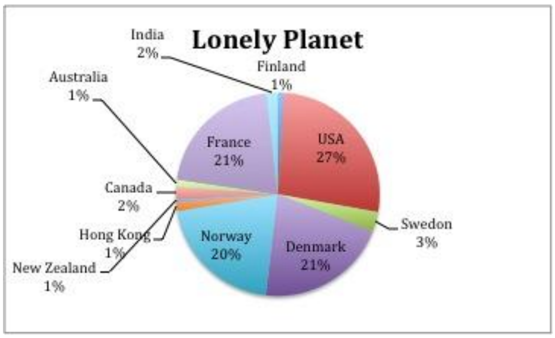


Figure 18: Popularity share of Lonely Planet on twitter

The pie chart of Lonely Planet shows a popularity share of 27% for USA followed by France (21%), Denmark(21%),, Norway(20%). Rest all countries doesn't discuss much about Lonely Planet on twitter.

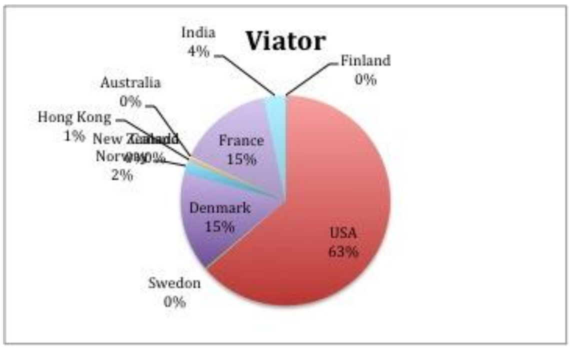
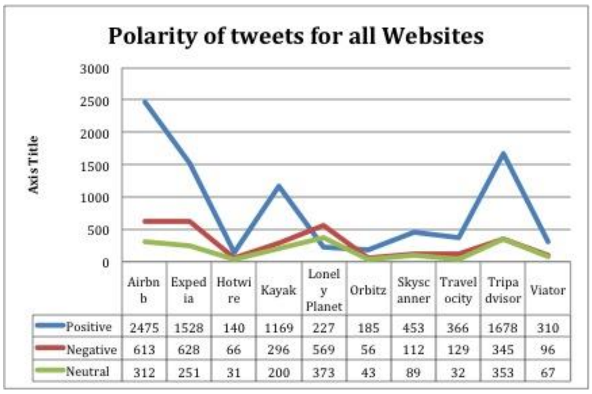


Figure 19: Popularity share of viator on twitter

The above chart depicts the popularity share of Viator on twitter. USA captures the most tweets for viator with a 63% popularity share.



	Airbnb	Expedia	Hotwire	Kayak	Lonely Planet	Orbitz	Skyscanner	Travelocity	Tripadvisor	Viator
Positive	2475	1528	140	1169	227	185	453	366	1678	310
Negative	613	628	66	296	569	56	112	129	345	96
Neutral	312	251	31	200	373	43	89	32	353	67

Figure 20: Polarity of travel websites

The above figure is a representation of the polarity of websites. This polarity graph shows that Airbnb has the positive tweets with 73% of the share. Airbnb dominates being 2% ahead of expedia which holds 71% of the positive tweets. However, Kayak is right behind tripadvisor with a minor 1% gap at 70%. Skyscanner and travelocity, at 69%, are the next in holding a considerable share of positive tweets in this representation. Viator(66% positive tweets), Orbitz(65% positive tweets) and Expedia(64% positive tweets) lag behind their competitors. Lonely Planet with only 19% positive tweets don't seem to compete with these big players This pie chart gives us a clear picture of the most positive tweets market players. While the competition seems difficult of the face of it, there is a huge gap of positive tweets between Airbnb and Lonely Planet.

CONCLUSION

Travel industry is one of the biggest service industries in the world. Our research provides a platform for billions of travellers to study fellow travellers sentiments and opinion about travel companies. The research equips travellers to an informed decision about their favored travel website. This is a leap forward in travel industry because it prevents people from getting misled during their travel. The paper concludes that Airbnb is the most trusted company in the world, followed by Tripadvisor and Skyscanner. Lonely planet falls last under this category with least number of positive tweets around the globe. The most talked about company in the world is Airbnb followed by expedia and tripadvisor. The least talked about company is Hotwire. The research will help the travellers to save time and energy, make their searches efficient and focused. The research helps the travel industry to self analyze their processes and products. The companies will be able to know about their shortcoming in terms of location, polarity and sentiments.

FUTURE SCOPE

The paper opens a new way of analyzing tweets in travel industry. There is a large scope of making new discoveries in this field. The work can be taken further by analyzing sentiments and opinions on various travel companies at various other social media platforms. This will help reveal different aspects of travel industry.

REFERENCES

[1] M. d. P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. Á. Rodríguez-García, and R. Valencia-García, "Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach," *Computational and mathematical methods in medicine,* vol. 2017, 2017.

[2] A. Kanavos, N. Nodarakis, S. Sioutas, A. Tsakalidis, D. Tsolis, and G. Tzimas, "Large Scale Implementations for Twitter Sentiment Classification," *Algorithms,* vol. 10, no. 1, p. 33, 2017.

[3] M. M. Altawaier and S. Tiun, "Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis," *International Journal on Advanced Science, Engineering and Information Technology,* vol. 6, no. 6, pp. 1067-1073, 2016.

[4] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques," *Journal of King Saud University-Computer and Information Sciences,* vol. 28, no. 3, pp. 330-344, 2016.

[5] M. Sodanil, "Multi-Language Sentiment Analysis for Hotel Reviews," in *MATEC Web of Conferences,* 2016, vol. 75, p. 03002: EDP Sciences.

[6] J. Bian *et al.*, "Mining twitter to Assess the public perception of the “Internet of Things”," *PloS one,* vol. 11, no. 7, p. e0158450, 2016.

[7] E. M. Clark *et al.*, "Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter," *PloS one,* vol. 11, no. 7, p. e0157304, 2016.

[8] W. Hua, Z. Wang, H. Wang, K. Zheng, and X. Zhou, "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge," *IEEE transactions on Knowledge and data Engineering,* vol. 29, no. 3, pp. 499-512, 2017.

[9] S. Caton, M. Hall, and C. Weinhardt, "How do politicians use Facebook? An applied Social Observatory," *Big Data & Society,* vol. 2, no. 2, p. 2053951715612822, 2015.

[10] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PloS one,* vol. 10, no. 12, p. e0144296, 2015.

[11] E. M. Cody, A. J. Reagan, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Climate change sentiment on twitter: an unsolicited public opinion poll," *PloS one,* vol. 10, no. 8, p. e0136092, 2015.

[12] C. Clavel and Z. Callejas, "Sentiment analysis: from opinion mining to human-agent interaction," *IEEE Transactions on affective computing,* vol. 7, no. 1, pp. 74-93, 2016.

[13] K. Holmberg, T. D. Bowman, S. Haustein, and I. Peters, "Astrophysicists’ conversational connections on Twitter," *PloS one,* vol. 9, no. 8, p. e106086, 2014.

[14] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal,* vol. 5, no. 4, pp. 1093-1113, 2014.

[15] T. Desai *et al.*, "Tweeting the meeting: an in-depth analysis of Twitter activity at Kidney Week 2011," *PloS one,* vol. 7, no. 7, p. e40253, 2012.

[16] C. Chew and G. Eysenbach, "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak," *PloS one,* vol. 5, no. 11, p. e14118, 2010.

[17] P. Bhardwaj, , S. Gautam, , P. Pahwa, “A Novel Approach to Analyze the Sentiments of Tweets Related to Tripadvisor”, unpublished

[18] “The world's biggest travellers - which nationality takes the most trips?” ,http://www.traveller.com.au/the-worlds-biggest-travellers--which-nationality-takes-the-most-trips-11979j