Netnography and Text Mining to Understand Perceptions of Indian Travellers using Online Travel Services

Dashrath Mane[1], Dr. Prateek Srivastava[2]

Department of CSE, Sir Padampat Singhania University [SPSU]

Udaipur, India

Abstract—Advancements in the electronic commerce industry have helped online travel services (OTA) in many ways. The paper examines the overall impact of traveller's using online services and their sentiments derived from a collection of reviews for online travel service providers known as online travel agents (OTA) in India. Customer reviews from different identified sources are collected and the satisfaction of travellers using various online travel services is analyzed using netnographic analysis and text mining. This paper also covers a detailed process of data collection, analysis using netnography and text mining methods which helps us for the analysis and deriving sentiments from collected reviews. Various results obtained are presented as part of token lists, keyword analysis, and service-specific analysis. The statistical analysis of different results is tested to understand the relationship between various services and OTA.

Keywords—Consumer; travellers; netnography; text mining; OTA; sentiment; perception

I. INTRODUCTION

A. Tourism Market in India

India, the country is known for its rich history, cultures. India placed 40th out of 136 nations as per the report of Travel & Tourism Competitiveness for the year 2017 [1]. The tourism industry in India is playing a significant role in the economy [2].

The current scenario shows that India has a bright future in the tourism sector as lots of development taking place in the Indian tourism industry. In the last few years, the Ministry of Tourism, Government of India, has initiated various attempts to lift tourism in a country that has opened research opportunities in it. The plans like PRASAD, e-tourist Visa, Mobile applications for tourists are helping out to gain a new tourist footprint [3].

B. Online travel services in India

It is expected that by the end of 2020, the count of the bookings made online is likely to reach 4 billion in India.[4] Makemytrip.com has been positioning high on the rundown of movement sites that fill in as a one-stop look for the whole group. The online booking travel related firm is a commonly known name these days & it has been collaborating with imperative players in the industry. Yatra.com[5] is the best for movement specialists, long-standing corporate customers, and easygoing hikers alike & is an excellent travel site in India. The best costs and cashback bargains on air are assured with yatra.com.

Cleartrip online travel service provider offers internet booking for train and flight tickets for domestic and international bookings. Goibibo.com is another name which offers the best arrangements/bookings in trains, flight. From booking universal/local occasions to encouraging Foreign Exchange, Visa, Passport, protection for movement purposes. Thomas cook does this everything for the travellers. Travelguru.com, regularly appraised as a standout amongst the most went by Indian travel destinations on the web, is a fortune place of treats with regards to booking air tickets, lodging offices.

Netnography and Tourism industry: Netnography is Method used for studying culture and communities online. It is a tool to understand social interaction in digital communications [6]. Netnography is a known tool accepted as a virtual ethnography tool, is being explored along with text mining. The approach used in this paper also produces an analysis of perceptions of travellers in India using text mining techniques.

Phases in the Netnography Process [11]: Fig. 1 explains various phases in the process of netnographic analysis; below are the various phases in it:

1) Research planning: Research planning [3] is the initial phase in netnographic studies. This phase speaks about defining the problem, defining the research objectives, should talk about translating the research objectives into a specific set of questions.

2) Entrée: In this phase, online communities, blogs, groups are identified and allows an ethnographic to enter into the communities to get a better knowledge of cultures and communities.

3) Data collection: Data collection is a very vital phase in this type of study. In this phase, Netnographer collects data from Internet data, interview data, and field notes. Travellers' reviews downloaded using various rating & referral sites like mouthshut.com and consumeraffairs.com. These are the websites that allow the user to collect millions of reviews posted by travellers in India. The whole reviews were collected using systematics steps in the netnographic process[7].

4) Interpretations/Data analysis: This makes use of One of the data analysis technique known as Analytical coding, Renal data analysis consist of coding, noting, abstracting, checking and refining, generalizing and **t**heorizing as shown in Fig. 2.

5) The data collection process was performed from sources like mouthshut.com [8] and Consumer Affairs.com for this study. More than 2000 reviews from these platforms have been collected and maintained in an excel sheet.
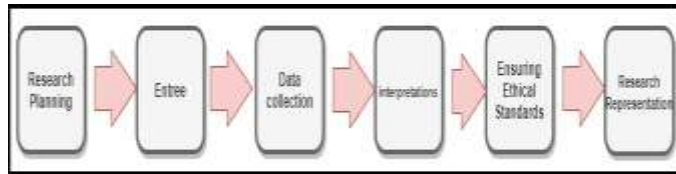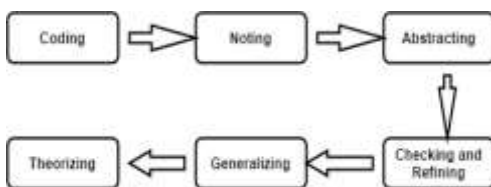

Fig. 1. Phases in Netnography Process [11].

Fig. 2. Analytical Data Analysis [11].

II. RESEARCH METHODOLOGY

The online travel agents considered in this research work includes MakeMyTrip, Goibibo, Cleartrip, redBus, and Yatra. A good number of reviews for each of these agents [12] selected are evaluated using netnographic studies. The analysis performed as netnographic analysis is used to carry out the sentiments, the levels of sentiments of travellers using this online travel-related online services in India. Fig. 3 shows the overall rating for Cleartrip on mouthshut.com.

Fig. 4 represents the sample to review and rating for Cleartrip [7]. In addition to actual comment, Reviews contain values from 1 (low) to 5(high) for all the website components.

Fig. 5 is a sample of the overall rating for MakeMyTrip. Referring to the Fig. 5, it is very much clear that for all five website components, the overall rating for MakeMyTrip is 2 [8].

Data collected from the identified referral, rating sites is stored in the Excel sheet. This data contains various fields as 1) Date of review, 2) Reviewer name, 3) Gender, 4) Age, 5) Location, 6) Review, 7) Source, 8) Review rating, 9) Service and support, 10) Information depth, 11) Content, 12) User-friendly, 13) Time to load.

Fig. 6 shows an Excel document of reviews collected from different sources of online travel agents.

A. Text Pre-Processing

The reviews were collected from various online platforms & referral rating sites need to have a series of text preprocessing before referring them for analysis using Text mining. The concept of text preprocessing consists of a series of stages, which include spelling in normalization, filtering, lemmatization. The various text preprocessing tasks are being essential and this includes content cleanup, tokenization, grammatical form tagging [9].

Table I represent various phases in determining positive negative reviews the process of determining sentiments consist of some of the essential processes like splitting words, POS tagging, lemmatization, joining and then sentiment analysis.
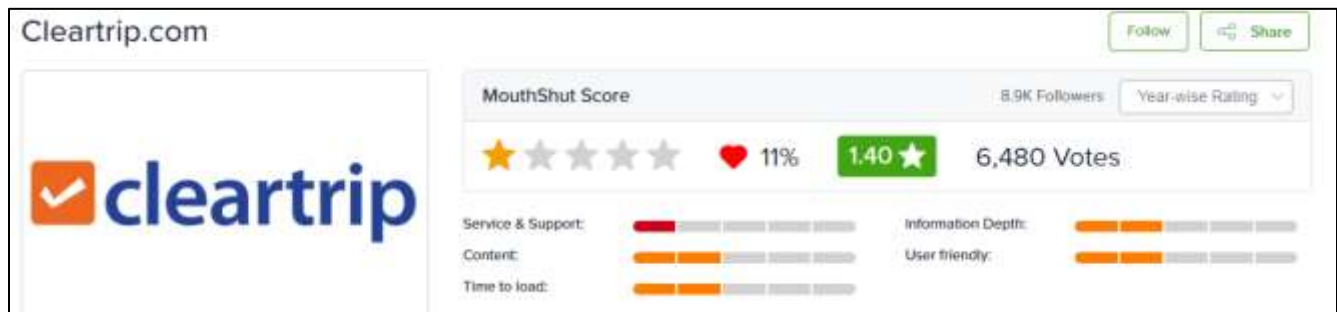

Fig. 3. The Overall Rating for Cleartrip (Source: Mouthshut.com).

Fig. 4. Sample Review and Rating for Cleartrip (Source: Mouthshut.com).

Fig. 5.   Overall Rating for MMT (Source: Mouthshut.com).

Fig. 6.   Different Fields in Format of Reviews Collected.

TABLE I.   VARIOUS PHASES IN DETERMINING POSITIVE-NEGATIVE REVIEWS

Process	Output
Initial Review	Horrible experience, hotels don't provide basic amenities and charge for everything. Customer service part is worst.
Lowercase	Horrible experience, hotels don't provide basic amenities and charge for everything. Customer service part is worst.
Punctuation	Horrible experience hotels don't provide basic amenities and charge for everything customer service part is worst
Split words	['horrible', 'experience', 'hotels', "don't", 'provide', 'basic', 'amenities', 'and', 'charge', 'for', 'everything', 'customer', 'service', 'part', 'is', 'worst']
POS tagging	[('horrible', 'JJ'), ('experience', 'NN'), ('hotels', 'NNS'), ("don't", 'VBP'), ('provide', 'VB'), ('basic', 'JJ'), ('amenities', 'NNS'), ('and', 'CC'), ('charge', 'NN'), ('for', 'IN'), ('everything', 'NN'), ('customer', 'NN'), ('service', 'NN'), ('part', 'NN'), ('is', 'VBZ'), ('worst', 'JJ')]
Lemmatization	['horrible', 'experience', 'hotel', "don't", 'provide', 'basic', 'amenity', 'and', 'charge', 'for', 'everything', 'customer', 'service', 'part', 'be', 'bad']
Joining	horrible experience hotel don't provide basic amenity and charge for everything customer service part be bad
Sentiment Analysis	{'negative': 0.333, 'neutral': 0.667, 'positive': 0.0, 'compound': -0.7906}
Sentiment	Positive / Negative

III. RESULTS AND DISCUSSIONS

The results obtained from this research are presented in this section.

1) Token identification: Various subsections are being identified based on the results. After processing the textual reviews collected using text mining techniques following are the detailed list of tokens identified. The below section represents the top 25 tokens from the overall volume of reviews processed and also online agent-specific [10].

Tokens identified are used for deriving the overall perception of the travellers for the overall. Table II represents the top 25 tokens from both the categories positive and negative. The top 25 tokens based on the frequency of their occurrence has shown in the Table II For overall positive tokens care, support, help, these are the frequently used words by the reviewers while posting the reviews. Similarly, in the negative tokens category words like problems, fraud, cheat, and mistakes are some of the commonly used words by the reviews while expressing on social platforms.

Fig. 7 and 8 graphically represents the top 5 tokens in positive and negative category respectively

2) Gender-specific (Male) Analysis of tokens: In the gender-specific category of tokens obtained, the Table III represents the top tokens based on their frequency of occurrence in reviews that are listed. It is also identified from the Table III that when it comes to male support, care, friend, help these are the frequently used words while expressing in the form of review. And it denotes the satisfaction of the travellers the category of male.

Similarly, for or negative categories in males, the commonly used words based on the results obtained are a problem, fraud, Waste, mistake. This represents the dissatisfaction of travellers regarding online travel-related services. Tables III and IV represents the top 5 positive and top 5 negative tokens for males. Overall male analysis performed and results obtained indicate that usually while expressing on social platforms about satisfaction, travellers have used quite similar words.

TABLE II.   OVERALL POSITIVE-NEGATIVE TOKENS

Overall Positive Tokens					Overall Negative Tokens			
Sr. No	Words	Frequency	Polarity		Sr. No	Words	Frequency	Polarity
1	Care	489	0.49		1	Problem	244	-0.4019
2	Support	454	0.40		2	Fraud	153	-0.5859
3	Friend	267	0.49		3	Cheat	70	-0.4588
4	Please	191	0.32		4	Mistake	65	-0.34
5	Help	189	0.40		5	Cheater	60	-0.5423
6	holiday	133	0.40		6	Fault	57	-0.4019
7	thanks	109	0.44		7	Waste	50	-0.4215
8	credit	90	0.38		8	Emergency	47	-0.3818
9	Kind	88	0.53		9	Delay	45	-0.3182
10	trust	65	0.51		10	Error	35	-0.4019
11	thank	59	0.36		11	Loss	33	-0.3182
12	solution	42	0.32		12	Complain	27	-0.3612
13	promise	40	0.32		13	Trouble	25	-0.4019
14	value	37	0.34		14	Penalty	25	-0.4588
15	hope	30	0.44		15	Shock	22	-0.3818
16	comfort	29	0.36		16	Hell	21	-0.6808
17	profit	26	0.44		17	Fool	20	-0.4404
18	resolve	25	0.38		18	inconvenience	19	-0.3612
19	hand	23	0.49		19	Doubt	19	-0.3612
20	satisfaction	22	0.44		20	Rude	16	-0.4588

Fig. 7.   Overall Positive Tokens (Top Five).

Fig. 8.   Overall Negative Tokens (Top Five).

Overall male analysis performed and results obtained indicate that usually while expressing on social platforms about dissatisfaction, travellers have used quite similar words.
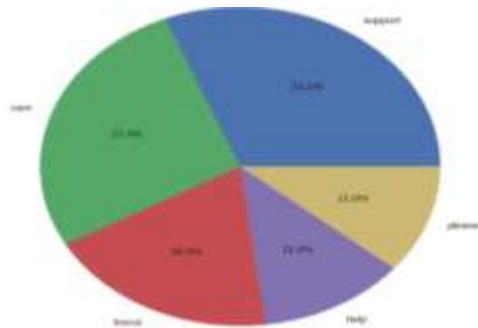
TABLE III.   TOP FIVE POSITIVE MALE

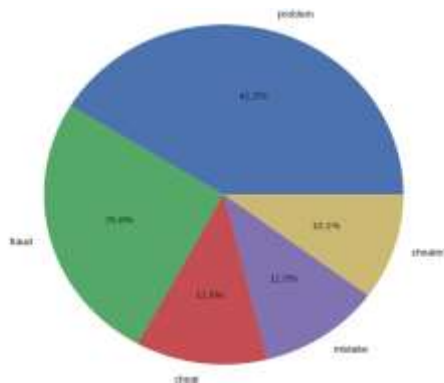| Overall Male Positive			
Sr. No	Words	Frequency	polarity
1	Support	327	0.40
2	Care	284	0.49
3	Friend	199	0.49
4	Help	125	0.40
5	please	116	0.32

TABLE IV.   TOP FIVE NEGATIVE MALE

Overall Male Negative			
Sr. No	words	Frequency	Polarity
1	problem	173	-0.4019
2	Fraud	89	-0.5859
3	Waste	34	-0.4215
4	mistake	33	-0.34
5	emergency	33	-0.3818

3) Gender-specific (Female) Analysis of tokens: This section describes various tokens used in the form of reviews while expressing on social platforms by female travellers. The analysis carried out in below Table V and VI represents the top 5 tokens based on their value of frequency. It also represents their importance while deriving positive and negative sentiments from the reviews hosted by female travellers. In a positive category, the frequently used word by female Travellers is Care, support, please, and so on. In negative categories, the most frequently identified words are problem fraud, cheat, a mistake.

Tables V and VI are the tabular representation of the top 5 reviews from gender-specific analysis of tokens in female traveller.

Overall female analysis performed and results obtained indicate that usually while expressing on social platforms about dissatisfaction, travellers have used quite similar words.

The analysis carried out in table VI represents the top 5 tokens based on their value of frequency for the positive female category.

Table VII represents a comparison matrix with the top ten tokens in each OTA. Based on the results shown, it is very much clear that satisfied travellers are frequently expressing with words like support, help, kind, comfort, and dissatisfied travellers are using words like problem, fraud, cheater, mistake while expressing on social platforms.

The comparison matrix shows the frequency of particular words when Traveller post their comments reviews on various sites. This comparison matrix also helps us understand comparative analysis between the keywords in the form of tokens for all the five online travel agents chosen in this research. Words like problem, help, and support are common in all the OTA.

4) Keyword Analysis: Following part of the research paper presents various keyword visualizations to understand more importance of each of the keyword plays in the entire study. Fig. 9 explains lines of code referred from Python-based implementation which has helped to get the overall graphical representation of keywords and its analysis.

Fig. 10 represents the first Visualization of keywords for the real data values collected for the research purpose. From the results obtained, it very much clear that words like Hotel, book, ticket, service, call are the most critical words based on their appearances.

Fig. 11 and 12 represents the essential keywords in positive and negative category for the entire study. From the overall positive category, it is very much clear that words like best, awesome, trust, kind, super are the critical words describing positive sentiments in this category. In contrast, in overall negative word analysis. Words like fraud, bad, horrible, pathetic, failed is the crucial words that frequently occurred while expressing sentiments by the Travellers.

TABLE V. TOP FIVE POSITIVE FEMALE

Overall Female Positive			
Sr. No	Words	Frequency	polarity
1	Care	193	0.49
2	support	115	0.40
3	please	73	0.32
4	friend	63	0.49
5	Help	62	0.40

TABLE VI. TOP FIVE POSITIVE FEMALE

Overall Female Negative			
Sr. No	words	Frequency	Polarity
1	problem	68	-0.4019
2	Fraud	60	-0.5859
3	Cheat	34	-0.4588
4	mistake	31	-0.34
5	cheater	29	-0.5423

TABLE VII. COMPARISON MATRIX FOR OTA'S AGAINST TOP 10 TOKENS

Sr. No	Yatra		redBus		MMT		Goibibo		Cleartrip	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
1	care	problem	Care	problem	support	problem	care	problem	support	fraud
2	please	fraud	Support	fraud	care	cheat	support	fraud	care	cheater
3	support	waste	Friend	mistake	friend	fraud	friend	cheat	friend	problem
4	holiday	cheat	Please	delay	holiday	mistake	please	cheater	help	cheat
5	help	emergency	Help	emergency	help	error	help	waste	please	fault
6	thanks	mistake	kind	fault	please	emergency	thanks	fault	credit	delay
7	friend	penalty	thanks	waste	kind	loss	credit	mistake	thanks	mistake
8	credit	fraudsters	comfort	trouble	thanks	fault	holiday	fake	thank	cheating
9	kind	Cheater	thank	cheater	promise	penalty	trust	loss	trust	doubt
10	value	Delay	trust	rude	credit	cheater	kind	fool	holiday	scam

```
wordcloud = WordCloud( background_color ='white')

wordcloud.generate_from_frequencies(frequencies=d1)

plt.figure()

plt.imshow(wordcloud, interpolation="bilinear")

plt.axis("off")

plt.savefig(abc+'_overall.png', dpi=600)

plt.show()
```

Fig. 9.   Keyword Analysis LOC.

Fig. 10. Visualization of Overall Word Cloud.

Fig. 11. Overall Satisfied Travellers.

Fig. 12. Overall Dissatisfied Travellers.

The section below describes the online travel agent (OTA) Specific keyword analysis.

5) Cleartrip Keyword analysis: Fig. 13 and 14 depicts the visualization of keywords in Cleartrip. Based on the results obtained, it is very much clear that words like best, trust, great, happy are the most critical in a positive category for Cleartrip. In contrast, Fig. 13 shows bad, fraud, pathetic, fail other common words in negative categories for Cleartrip.

Fig. 13. Cleartrip Satisfied Travellers Word Cloud.

Fig. 14. Cleartrip Dissatisfied Travellers Word Cloud.

6) Goibibo Keyword analysis: Fig. 15 and 16, covers keywords based on their frequent occurrence for Goibibo. Based on the analysis carried out for Goibibo keywords excellent, best, great, trust, free are the crucial words in the positive category for Goibibo. Whereas worst, bad, fraud, hate is the commonly observed keywords in the negative category for Goibibo.

7) MMT Keyword Analysis: MakeMyTrip (MMT) keyword analysis is expressed in Fig. 17 and 18 in graphical representation. In a positive category, best, free, kind, happy, wonderful are critical words for MMT. In contrast, when it comes to negative category pathetic, bad, worst, horrible, fraud, these are the words carrying much importance in MMT keyword analysis.

Fig. 15. Goibibo Satisfied Travellers Word Cloud.

Fig. 16. Goibibo Dissatisfied Travellers Word Cloud.

Fig. 17. MMT Satisfied Travellers Word Cloud.

Fig. 18. MMT Dissatisfied Travellers Word Cloud.

*8) redBus Keyword Analysis***:** Similar to other online travel agents, keyword analysis is performed for redBus. Fig. 19 and 20 depicts various important words in positive and negative categories for redBus. Best, happy, kind, comfortable these are the top-rated words in the positive category, and bad, worst, horrible these are some of the highly-rated words in the negative category for redBus.

9) Yatra Keyword Analysis: Fig. 21 and 22 is the word count analysis done on online reviews collected for Yatra online travel agent. The outcome of the analysis demonstrates the words like best, great, kind, splendid, love is the essential words in the positive category for the Yatra. In the negative category, analysis says the words like bad, fraud, worst, unprofessional, horrible are some of the words that play an essential role in Yatra.

Fig. 19. redBus Satisfied Travellers Word Cloud.

Fig. 20. redBus Dissatisfied Travellers Word Cloud.

Fig. 21. Yatra Satisfied Travellers Word Cloud.

Fig. 22. Yatra Dissatisfied Travellers Word Cloud.

10)Analysis based on Rating: Each of the review collected from the Mouthshut.com has a rating value attached to it ranging from 1 as Low to 5 as High. Following is the OTA's and reviews based on rating value percentage Fig. 23 covers lines of code for getting the statistical value of review ratings for all the reviews which are being collected and processed and relevant graphical representations are obtained with the help of similar lines of code.

The Table VIII represents a statistical analysis of reviews regarding the overall rating given by the reviewer on the scale of 1 to 5, where 1 is low, and 5 is high. The table shows a good number of reviews for each of the online travel agents as review rating low, which means dissatisfaction of travellers is low. At the same time the volume of reviews classified under 5 represents the high level of satisfaction of travellers using an online travel agent and there relevant services MakeMyTrip, Goibibo and Cleartrip are having good volume of reviews under the highest rating.

```
sum = []
percent = []
for i in range(5):
    sumval = 0
    for j in range(5):
        sumval += len(ratingList[j][i])
    sum.append(sumval)
for i in range(5):
    percent = []
    for j in range(5):
        percent.append(round(((len(ratingList[j][i]) * 100) /
sum[i] ) , 1 ))
        worksheet.write(i, j, round(((len(ratingList[j][i]) * 100) /
sum[i] ) , 1 ))
    worksheet.write(i, 5, sum[i])
    percent.append(sum[i])
    print(percent)
workbook.close()
```

Fig. 23. Rating based Analysis LOC.

TABLE VIII. OTA'S REVIEW RATING

Review Rating		1	2	3	4	5	Total Reviews
OTA's	Cleartrip	236	13	12	33	57	351
	Goibibo	321	33	36	54	75	519
	MMT	281	18	48	90	94	531
	redBus	292	38	45	46	48	469
	Yatra	172	17	15	40	21	265

Table IX represents the statistical analysis of review ratings for all the OTA in percentage collected and processed for each of these online travel agents. From the table, it is very much clear that MakeMyTrip is having the highest percentage of review rating value with 5. It means the satisfaction level of travellers using various services offered by MakeMyTrip is also high in comparison to other online travel agents.

Cleartrip has 16.2% reviews based on the overall rating given to the review this percentage and it's second after MakeMyTrip. Goibibo is third in the list with the highest percentage under 5. It is very much clear from table values that Yatra is having low preference under the top-level, which means satisfaction is very low for the Yatra.

TABLE IX. OTA'S REVIEW RATING PERCENTAGE

Review Rating		1	2	3	4	5	Total volume
OTA	Cleartrip	67.2	3.7	3.4	9.4	16.2	351
	Goibibo	61.8	6.4	6.9	10.4	14.5	519
	MMT	52.9	3.4	9	16.9	17.7	531
	redBus	62.3	8.1	9.6	9.8	10.2	469
	Yatra	64.9	6.4	5.7	15.1	7.9	265

Looking at Table IX for Yatra, 64.9 percent of Travellers have given Low review rating, represent the percentage of the satisfaction of Travellers using services offered by Yatra.

IV. ARCHITECTURAL VIEW OF MODEL

In this research work, netnography and Text mining techniques are used as an integrated approach. The following are the various stages in the architectural view of the basic model in this study.

Fig. 24 represents the architectural view of the platform used in the text mining process. As mentioned in the diagram, the following are the steps that help get a visualized representation of specific results. The first stage in the architectural view is inputting of the review. In this research study, the reviews or posts used are collected from various online platforms using Netnographic guidelines. Collected reviews are maintained in a repository. The second stage is the preprocessing stage. In this stage, preprocessing of the text reviews collected is done using various techniques of splitting the words, converting the words, checking for the missing values is performed. The next stage is cleaning the reviews; in this process of cleaning of the reviews, it is very much essential that some of the processes which help the text to

refine and get a meaningful text for further processing. Sentiment score stage, the sentiments of each word, and then for the entire review are obtained.in this stage, using the Vader sentiment lexicon dictionary. The sentiments are derived as positive, negative, neutral—compound sentiment score help to classify the review into positive or negative. In the review classification stage, based on the compound sentiment scores with more than 0.05 compounded courses are classified as positive, and the sentiment scores between -0.05 and 0.0 are considered neutral review. The review has a compounded score of less than -0.05 classified as negative reviews. In the visualization stage, based on the review, the various analysis has been carried out in the graphical representation has been generated in the form of the word cloud, pie charts.
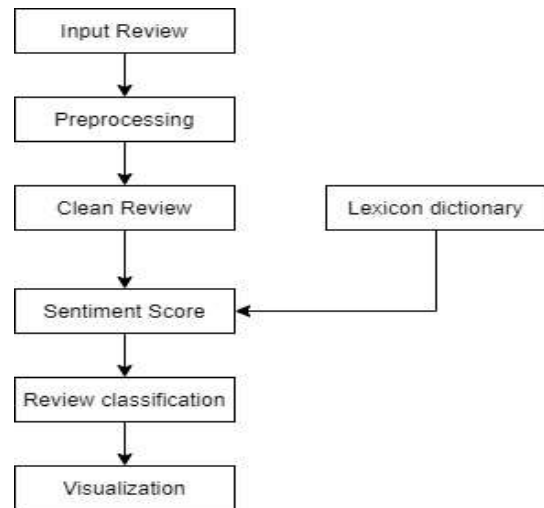

Fig. 24. Architectural view of Approach used for Visualization.

V. STATISTICAL ANALYSIS OF RESULT

In this section, the statistical analysis results obtained using the Chi-square test are discussed. A Chi-square test for various parameters is performed to understand and test the relationship between various variables based on the results of the test acceptance or rejection of the hypothesis is being chosen. Performed Chi-square test on the following scenarios:

1) Scenario 1(OTA vs. Gender): Here, the Chi-square statistical test is used to test whether two variables are independent or not. Chi-square hypothesis testing is used to understand whether there is any relationship between the online travel agent and the gender of travellers.

a) Null hypothesis (H0): No association/relationship between gender of traveller and OTA.

b) The alternate hypothesis (H1): Gender of traveller and OTA has an association/relationship

Table X(a) and X(b) shows the calculations of test statistics.

P-value obtained is .00000009307. With the degree of freedom 4 and at 5% level of significance, the critical value/tabular value is 9.49. The calculated chi-square value(33.54) is greater than the critical or table value(9.49).

TABLE X. (A) TEST STATISTICS FOR OTA VS. GENDER

Gender OTA	Male			Female			Total	
	Observed	Expected	%	Observed	Expected	%	Observed	%
Cleartrip	242	233.38	16.74	110	118.62	14.97	352	16.14
Goibibo	429	385.87	29.67	153	196.13	20.82	582	26.69
MMT	325	341.44	22.48	190	173.56	25.85	515	23.61
redBus	306	310.95	21.16	163	158.05	22.18	469	21.5
Yatra	144	174.37	9.96	119	88.63	16.19	263	12.06
Total	1446		100	735		100	2181	100

(b) Test Statistics for OTA vs. Gender

Observed	Expected	(obs-exp)^2	(obs-exp)^2/ Exp
242	233.38	74.30	0.32
429	385.87	1860.20	4.82
325	341.44	270.27	0.79
306	310.95	24.50	0.08
144	174.37	922.34	5.29
110	118.62	74.30	0.63
153	196.13	1860.20	9.48
190	173.56	270.27	1.56
163	158.05	24.50	0.16
119	88.63	922.34	10.41
Calculated Chi-Square Value			33.54

There is enough of the statistical evidence to reject the null hypothesis and to accept the fact that there is an association or relationship between various OTAs and the gender of travellers.

Since 33.52 > 9.48 or our P-value < 0.05

The alternate hypothesis is accepted. Thus there is an association between gender and people using OTA.

Fig. 25 is the graphical representation of observed values for males and females for all the online travel agents against the total volume of reviews collected.

2) OTA vs. Positive Negative Sentiments: In this section, the Chi-square statistical test performed to understand the relationship between positive negative sentiments of travellers and various online travel-related service providers (OTA).

a) Null hypothesis (H0): No association between Sentiment and people using OTA.

b) Alternate Hypothesis (H1): Association between Sentiment and people using OTA.

From Table XI, P-value obtained using the chi-square test is 0.01, chi-square value from P-value is 12.78; critical or tabulated chi-square value at the degree of freedom 4 and level of significance 5% is 9.49; since 12.78 > 9.48 An alternate hypothesis is accepted; it means an association or relationship between sentiments and people using OTA. Fig. 26 is a graphical representation of about table values representing the

relationship between OTA Sentiment against the total number of reviews for the online travel agents.

3) Statistical Analysis of Positive negative ratio Vs. Gender: In the below section, a detailed analysis of the relationship between positive-negative sentiment ratio and gender. Based on the data shown in the Table, P-value using the Chi-square statistical test is 0.00. The following are the null and alternate hypotheses framed to identify whether there is any relationship between the positive-negative sentiment ratio and gender.

a) Null hypothesis (H0): No association between Sentiment and Gender

b) Alternate Hypothesis (H1): Association between Sentiment and Gender

Referring to Table XII, The Chi-square value calculated using P-value is 12.96. And with a 5% level of significance and 4 degrees of freedom, the tabular value for chi-square is 9.48. Since the calculated value, 12.96 is more than the tabular value for chi-square 9.48. There is enough evidence to reject the null hypothesis and accept the alternate hypothesis, which says there is a relationship between Sentiment and gender. Since 12.95 > 9.48, Alternate hypothesis is accepted.

Fig. 27 represents a detailed classification of the total volume of reviews and sentiments as a positive and negative gender-wise. The finding from the analyses says that the male and female total percentage of negative review sentiment in male is less than female.
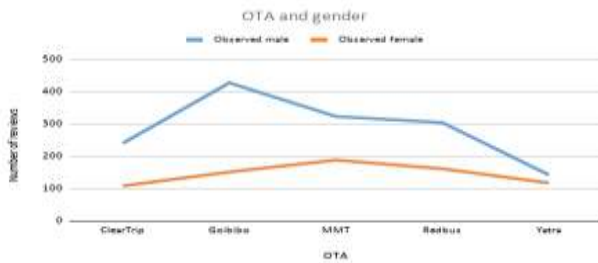
Fig. 25. OTA Specific Gender Values.

TABLE XI.   TEST STATISTICS FOR OTA VS. POSITIVE NEGATIVE SENTIMENTS

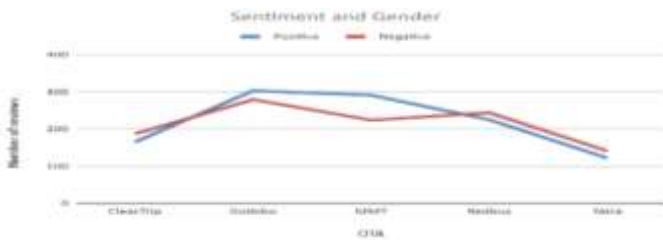| Gender OTA	Positive			Negative			Total
	Observed	Expected	%	Observed	Expected	%	Observed
Cleartrip	165	178.5	14.92	187	173.5	17.4	352
Goibibo	303	295.14	27.4	279	286.86	25.95	582
MMT	291	261.16	26.3	224	253.84	20.84	515
redBus	225	237.83	20.34	244	231.17	22.7	469
Yatra	122	133.37	11.03	141	129.63	13.12	263
Total	1106		100	1075		100	2181

Fig. 26. Sentiments and Gender Relationship for OTA.

TABLE XII.   RELATIONSHIP BETWEEN SENTIMENTS AND GENDER

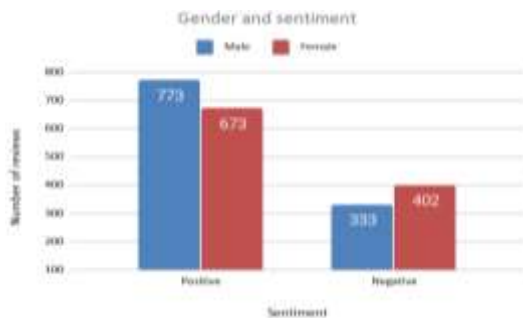| Gender Sentiment	Observed				Expected		Total
	Male	Male %	Female	Female %	Male	Female	
Positive	773	69.89	673	62.60	733.28	712.72	1446
Negative	333	30.11	402	37.40	372.72	362.28	735
Total	1106	100	1075	100			2181

Fig. 27. Classification Reviews and Sentiments.

VI. SERVICE SPECIFIC ANALYSIS

Online travel agents used in this research provides a various set of services to the users, customers, travellers. These services include hotel booking, flight booking, bus booking. In this section, the results obtained regarding these services are evaluated, and reasonable interpretations are derived. Fig. 28 shows lines of code help to do the analysis with reference to sentiments for each of the OTA specific to services identified.

```
xpos = np.arange(len(OTA))
plt.bar(xpos-0.2,posi, width=0.4, label="Positive")
plt.bar(xpos+0.2,negi, width=0.4,label="Negative")
plt.xticks(xpos,OTA)
plt.ylabel("Number of "+word+" reviews")
plt.title("Sentiment of "+word+" reviews")
plt.legend()
plt.savefig(word+"_OTA")
```

Fig. 28. Service Specific Sentiment Analysis LOC.

The graphical representation below Fig. 29 clearly indicates for bus-related bookings our services Travellers have preferred redBus.
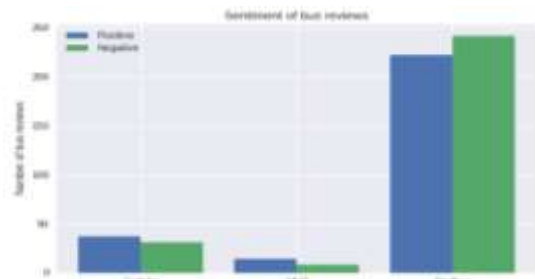

Fig. 29. Sentiments of Bus Reviews for OTA.

After redBus Travellers preferred Goibibo and then Makemytrip, the finding here is for bus-related services, redBus is on top of all.

Flight-related services are offered by the following online travel agents chosen for the study, Cleartrip, Goibibo, MakeMyTrip, and Yatra. The graphical representation Fig. 30 shows that for the flight-related services, the volume of negative reviews and sentiments are more with Cleartrip than other service providers. For Goibibo, the positive sentiments are little more than the negative sentiments which we can interpret in a manner that Travellers or customers are a little happier regarding flight-related services offered by Goibibo. Similarly, for MakeMyTrip, positive sentiments are more than negative sentiments for all the possible reviews processed for flight-related services which also represents the satisfaction of travellers is more than dissatisfaction for Makemytrip. For Yatra, the positive sentiments are less than the negative sentiments. Also, the volume of reviews for Yatra regarding flight-related services is low; it represents the satisfaction is

lower than the dissatisfaction and preference for flight-related services is not towards Yatra.
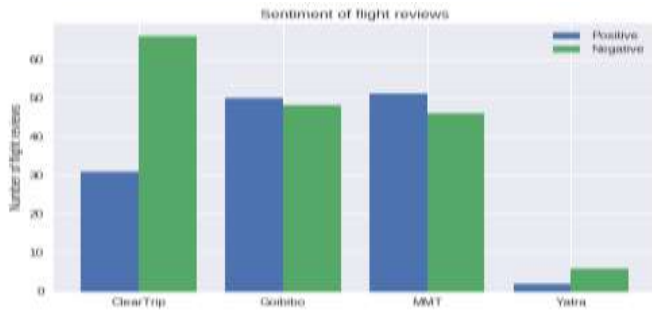

Fig. 30. The Sentiment of Flight Reviews for OTA's.

The following are the online travel agents that offer hotel booking related services online Cleartrip, Goibibo, MakeMyTrip, Yatra. Fig. 31 clearly explains the results of the analysis done regarding Hotel related services offered by these online travel agents in India. When it comes to Cleartrip Hotel related services, Fig. 31 represents the negative sentiments are more than positive it means when it comes to hotel-related services, online dissatisfaction is more observed in customers or travellers about Cleartrip. Goibibo looks at the top choice for Hotel related services online. Again based on the analysis, it is proved that positive sentiments are more than negative sentiments. It means the satisfaction of travellers using Hotel related services offered by Goibibo is high. MakeMyTrip also follows the line of Goibibo when it comes to hotel-related services. It has become the second preference for Hotel related services after Goibibo. It is also clearly indicated that Yatra is rarely preferred by travellers when it comes to online Hotel related services.

Fig. 31. The Sentiment of Hotel Reviews for OTA's.

VII. CONCLUSION

Customer is King for a successful business, customer satisfaction is considered to be a prime attribute. This research helps in consumer decision making while choosing various travel-related services and also businesses to choose their strategies to improve on services offered based on listening to the consumers.

Netnography and Text mining techniques used to perform analysis and processing of reviews and comments collected from various online platforms. The review collection has followed all the guidelines suggested by netnographic studies for five online travel agents i.e. MakeMyTrip (MMT), Yatra,

Cleartrip, Goibibo, redBus analysis is performed using text mining as an integrated approach.

Based on results obtained regarding service-specific analysis, it is possible to conclude that for flight booking, related services traveller's satisfaction is more with MMT, Goibibo. Consumers have preferred MakeMyTrip as the first option and Goibibo as Second and followed by Yatra. For Hotel booking related services satisfaction of travellers using services from MMT, Goibibo is high compared to others. For bus booking related services, users are more happy and satisfied with redBus then Goibibo and MMT.

This research work has used various approaches concerning the understanding of satisfaction, dissatisfaction, or the perception of travellers using selective travel-related online service providers in India to build a competitive advantage to customers and company too. Online travel service providers in India can make appropriate decisions according to the results obtained from the study regarding customer perception and create a competitive advantage. Also, consumers can choose a particular online travel-related service provider based on the result of the study for improved services and traveller satisfaction.

REFERENCES

[1] C. Science and M. Studies, "A Study on Tourist Satisfaction towards Hotel Related Services in Gujarat Tourism," vol. 4, no. 6, pp. 105–110, 2016.

[2] "Internet users in India to cross 500 mn in 2016: Prasad," May 2016.

[3] D. Mane and D. Bhatnagar, "Integrating Netnographic Analysis and Text Mining for Assessing Satisfaction of Travellers Visiting to India - A Review of Literature," 2020, pp. 564–572.

[4] "Top Online Travel Agencies (OTAs) in India - THE 'PERFECT' HOTEL." [Online]. Available: http://whiteskyhospitality.co.uk/top-online-travel-agencies-otas-in-india/. [Accessed: 28-Jul-2020].

[5] J. Chakravarthi and V. Gopal, "Comparison of traditional and online travel services: A concept note," IUP J. Bus. Strateg., vol. 9, no. 1, pp. 45–59, 2012.

[6] D. Mane., "Text Mining to Understand Major Keywords Explaining Sentiments of Travelers Using Travel Related Online," books.google.com.

[7] "CLEARTRIP.COM Reviews, Feedback, Complaint, Experience, Customer Care Number - MouthShut.com." [Online]. Available: https://www.mouthshut.com/websites/Cleartrip-com-reviews-925062909-srch. [Accessed: 28-Aug-2020].

[8] "MAKEMYTRIP.COM Reviews, Feedback, Complaint, Experience, Customer Care Number - MouthShut.com." [Online]. Available: https://www.mouthshut.com/websites/MakeMyTrip-com-reviews-925031929-srch. [Accessed: 28-Aug-2020].

[9] P. Gupta, R. Tiwari, and N. Robert, "Sentiment Analysis and Text Summarization of Online Reviews : A Survey," 2016 Int. Conf. Commun. Signal Process., pp. 241–245, 2016.

[10] Prameswari P., Surjandari I., and Laoh E.. Opinion Mining from Online Reviews in Bali Tourist Area. 226–230. (2017).

[11] Kozinets R., Netnography: doing ethnographic research online. Sage Publication, London (2010b).

[12] Top 10 travel Companies in India|TopEcommerceStartups., Retrieved from http://topecommercestartups.com/top-10-travel-companies-in-india/Travel& (2014).