

# Feature Based Opinion Mining using Text Analysis for Chhattisgarh Tourism Industry



Riya Sharma, Ani Thomas

**Abstract:** *Opinion mining is an approach of natural language processing (NLP) that distinguishes the emotional tone of the content or any sentence. This is often a well known approach to decide the assessment about an item, administration or thought. It includes the utilization of information mining, AI and man-made consciousness for conclusion and emotional data of the content. opinion mining is also referred as sentiment analysis .Sentient analysis can be said as study of human emotions .we can arrange those notions into positive and negative from any content . It is a procedure of evaluating the emotional value in content ,to have a comprehension of frames of mind, suppositions and emotions are expressed .The feedback of Tourists are important for Tourism Industries, because it enables to plan marketing strategies based on the reviews .So it is necessary to understand their sentiments about its distinctive features as overall sentiment of a place. Opinion mining is necessary for Chhattisgarh to enhance its tourism industry.*

**Keywords:** *Natural language processing, Opinion mining, Sentiment analysis, Tourism.*

## I. INTRODUCTION

Progression in web advances and correspondences affected the manner in which individuals can get to data. The web has become an enormous store of information, to which clients include new information online consistently. A piece of that data is spoken to by online surveys. Individuals currently read these surveys before buying or going somewhere and are affected by them during the time spent securing an item or administration. However, the huge measure of information makes unimaginable for one to peruse everything. So in this context it is increasingly critical to have a mechanized framework for gathering and preparing information, equipped for introducing to clients significant data an automated system for collecting and processing data, capable of presenting to users relevant information. Opinion mining can be used in various sectors such as tourism industry, e-commerce, banking, mining social media websites like Facebook, Twitter and so on. Using opinion mining and text mining, organizations can gain consumer reviews of a particular service from the response about them. This can be further used to study customer's satisfaction with the services and if in case complaints and issues, finding the possible reasons for that.

Websites, Forms, online journals, informal organizations, and substance sharing administrations help individuals to share their encounters, information and sentiments. Extraction of conclusions from online audits is a piece of another territory of research created in a decade ago.

Opinion mining considers the assurance and arrangement of assessments or emotions communicated in content, using registering machines. The challenge of the research area is to extract knowledge from unstructured data. The area of the travel industry broadened action online in the most recent decade.

There are many individuals that book settlement online in light of the fact that is less tedious, less expensive and they have the likelihood to get itemized data about offices and area of lodgings. Correspondingly to advancement of internet booking stages, destinations committed to introducing audits in the travel industry likewise developed. Booking destinations moreover fuse fragments with studies about presented lodgings.

The upside of moving toward information and information, make customers to lean toward web booking. Finds out about customers' online direct revealed that the decision of getting a thing is especially influenced by other buyers' appraisals.

A framework that could sum up the audits, separating the feelings from this data, offering a general viewpoint, would spare a great deal of time and facilitate the choice procedure for shoppers. Such a framework would likewise help supervisors to discover how their lodging is seen by clients, what administrations they loved or loathed. A productive criticism would help them on improving their administrations.

## II. THE PROPOSED SYSTEM ARCHITECTURE

The proposed system consists various steps:-

- i. Collection of data from various websites like trip adviser.
- ii. These datasets are to be cleaned by pre-processing.
- iii. Tokenization
- iv. Vectorization.
- v. Then we will extract different features according to words most occurring words related to the feature of places according to tourists review.
- vi. Then these reviews are classified.
- vii. After classification, the classified reviews are quantified as positive or negatives.
- viii. Finally the result is being interpreted.

Fig.1 shows the working of the proposed system

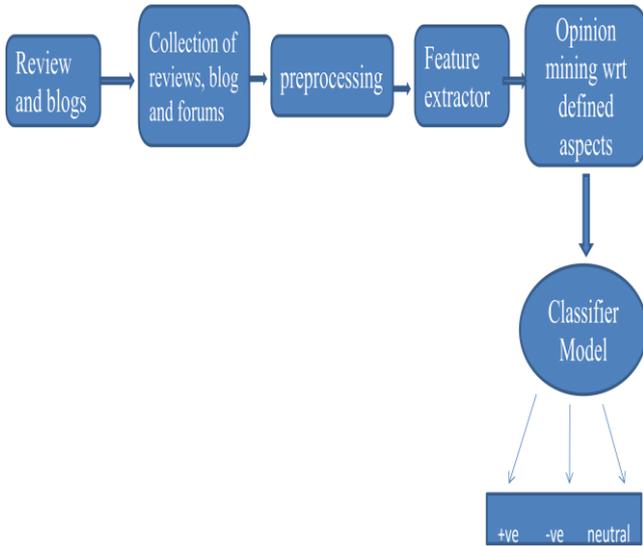
Revised Manuscript Received on July 02, 2020.

\* Correspondence Author

**Riya Sharma\***, Department of Information Technology, Bhilai Institute of Technology, Durg, Chhattisgarh, India. E-mail: riyas2304sharma@gmail.com

**Dr. (Mrs.) Ani Thomas**, Professor and Head, Department of Information Technology, Bhilai Institute of Technology, Durg, Chhattisgarh, India. E-mail- ani.thomas@bitdurg.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



**Fig.1**The proposed system architecture

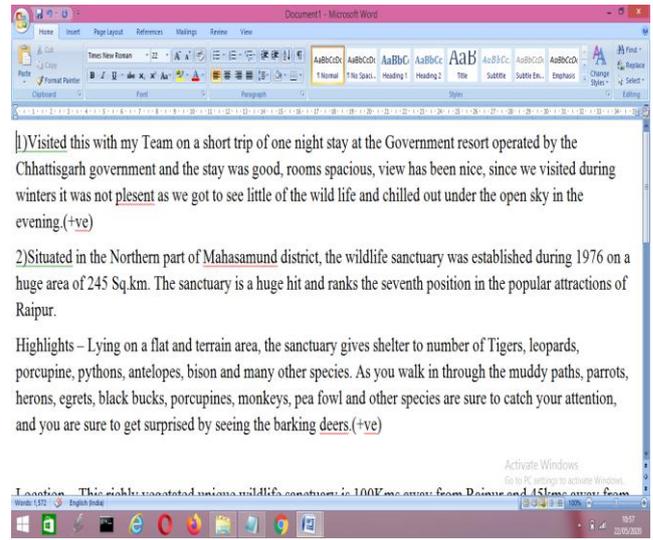
## A. Collection of dataset

Collection of tourists reviews are done from various websites like trip advisor, Twitter API. Reviews are taken for the places of Chhattisgarh like:- Chitrakoot, Gangrel Dam, Ghatarani, Jatmai, Jungle safari, Kailashkutumbsar cave, Barnawapara ,Kanger valley national park, Mm funcity, Muktangam, Rajim, Tirathgarh, Kanha national park etc.

Fig.2 shows the website from where data is being collected and Fig.3 shows the dataset.



**Fig.2:** data has been collected from official website of trip advisor



**Fig. 3:** dataset for reviews of Chhattisgarh tourists places

## B. Pre-processing

The database has to be cleaned so that we get proper text review on which analysis can be performed. Here we are to remove all unwanted characters like (!, ( , ) , (\*)) and emoticons etc which can disturb with our results. Fig.4 shows the dataset after preprocessing. Fig. 4 shows the codes for pre-processing.

```

1 import nltk
2 file1 = open("F:/chattisgarhreviews.txt", "rU")
3 raw = file1.readlines()
4 for f in raw:
5     asking = f
6     asking = ''.join([c for c in asking if c not in ('!', '#', '@', '?', ',')])
7     print asking
8     writing = open("F:/writechattisgarhreviews.txt", 'a')
9     writing.write('\n'+asking)
10    writing.close()
  
```

**Fig.4** pre-processed data

## C. Tokenization

We will tokenize all the cleaned information in our dataset. Tokens are singular terms or words and tokenization is the way toward parting a string of content into tokens. Tokenization is the demonstration of separating an arrangement of strings into pieces, for example, words, catchphrases, expressions, images and different components called tokens. Tokens can be singular words, states or even entire sentences. During the time spent tokenization, a few characters like accentuation marks are disposed of. The tokens become the contribution for another procedure like parsing and content mining. Fig.5 shows the tokenized data.

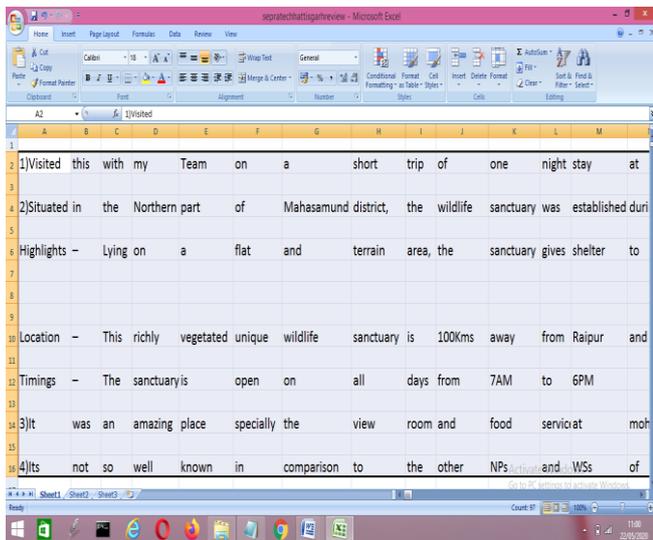


Fig.5: Tokenization

**D. Vectorization**

It is the transformation of words to vectors. Word2vec is a two-layer neural net that forms content. Its info is a book corpus and its yield is a lot of vectors: include vectors for words in that corpus. While Word2vec is certainly not a profound neural system, it transforms content into a numerical structure that profound nets can see Deep learning actualizes a circulated type of Word2vec for Java, which takes a shot at Spark with GPUs. Word2vec's applications reach out past parsing sentences in nature. It very well may be applied similarly also to genes, code, likes, playlists, internet based life diagrams and other verbal or representative arrangement in which examples might be recognized. Fig.6 shows the conversion of words to the vector.

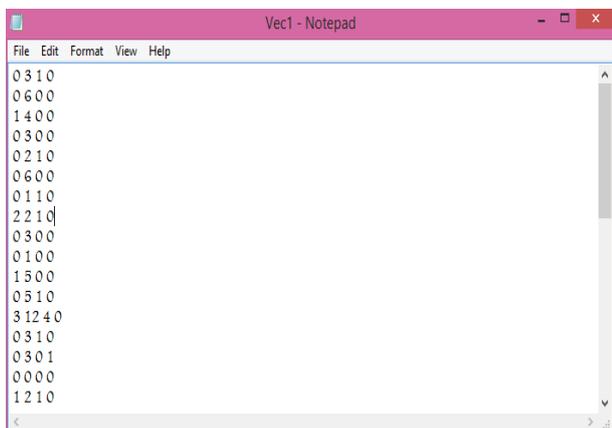


Fig.6: word to vector conversion

**E. Extracting Features**

Each review is parsed for the given feature. Only those reviews are collected whose words most closely match with the features provided. This is done by finding the feature words or its synonyms in the review. The word or its synonyms can be searched for in a sentence in a review and if found the review can be extracted else rejected. Each feature review will be stored in separate database file. Features like weather, food, visiting spots, cost etc are used for extracting features.

**F. Classifying**

Classification is at the core of both human and machine insight. Choosing what letter, word, or picture has been introduced to our faculties, perceiving countenances or voices, arranging mail, doling out evaluations to schoolwork, these are for the most part instances of relegating a class or classification to an info. We center around one normal content classification task, estimation examination, the ex-feeling investigation footing of opinion, the positive or negative direction that an essayist communicates toward some item. An audit of a film, book, or item on the web communicates the writer's opinion toward the item, while a publication or political content communicates assessment toward a competitor or politics. We can use different methods like support vector machine for proper classification.

**G. Result interpretation**

Finally we will Interpretate our result which shows that how much positive or negative a particular place's reviews are. Aspect based opinion mining for different tourist place of Chhattisgarh will be created.

**III. CONCLUSION AND FUTURE WORK**

We have seen that opinion mining can be utilized for breaking down sentiments in blog, articles, item surveys, social sites, film audit sites, the travel industry sites where a third individual portrays his perspectives. We likewise examined NLP and Machine Learning approaches for Sentiment Analysis.

We have analyzed Trip Advisor website and collected datasets. In those datasets we have performed pre-processing, then tokenization and then words are converted to vectors so that we can get better results. Further we can classify using support vector machine or naive bays classifier. We also studied NLP and Machine Learning approaches for Sentiment Analysis.

**ACKNOWLEDMENT**

This work is under Technical Education Quality Improvement Programme of Government of India (TEQIP) project.

**REFERENCES**

1. Cristian Bucur ,”Using Opinion Mining Techniques in Tourism”in 2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, 30-31 October 2014, Prague, Czech Republic.
2. S. Kavitha, S. Sathyavathi, S. Prabhakaran, S. Swathi” Opinion Mining on Tourism”in IJIRST –International Journal for Innovative Research in Science & Technology| Volume 3 | Issue 08 | January 2017 ISSN (online): 2349-6010
3. Smita Suresh Daniel, Dr. Neelam Sahu, Dr. Ani Thomas “An Algorithm to Quantify Sentiments of Product Reviews for a Given Feature using Text Mining”, Volume 7 Issue No.6, ISSN XXXX XXXX © 2017 IJESC.
4. Manisha Dubey, Prof.Tripti Sharma,”Ranking of Mobile Apps based on Opinion Mining” inInternational Journal of Innovative

- Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 5, May 2016
- Alaei, Ali Reza, Becken, Susanne, Stantic, Bela” Sentiment Analysis in Tourism: Capitalizing on Big Data” Journal of Travel Research, 2019, Griffith University Queensland Australia.
  - Girishkumar Sharma, promila sharma” A Study on Data mining Algorithms for tourism industry”, International Journal of Latest Trends in Engineering and Technology(IJTET), ISSN: 2278-621X..
  - Patel BhumiSomabhai, Tanvi Verma, Patel Pankita Somabhai “A Survey on Feature Based Opinion Mining for Tourism Industry”, Journal of Engineering Computers & Applied Sciences(JECAS) ,ISSN No: 2319-5606 , Volume 4, No.3, March 2015.
  - Ashutosh Bhatt & Ankit Patel. And Harsh Cheda,”Amazon Review Classification & sentiment analysis”, IJCSIT, 2015, vol6(6).
  - Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin ,”Sentimental analysis techniques in recent works” Science and Information Conference 2015 July 28-30, 2015 | London, UK.
  - Riya Sharma, Dr. Ani Thomas,” SENTIMENTAL ANALYSIS SURVEY ON RECENT TECHNIQUES”, International journal of management, technology and engineering Science, Volume IX, Issue III, MARCH/2019 ISSN NO : 2249-7455.
  - Smita Suresh Daniel, Dr. Ani Thomas, Riya Sharma, Niyati Motghare, Ashutosh Gupta,” SEMANTIC ANALYSIS FOR TEXT DATASET FROM AMAZON REVIEWS”, International Journal of Advanced in Management, Technology and Engineering Sciences, Volume 8, Issue III, MARCH/2018, ISSN NO : 2249-7455.

## AUTHORS PROFILE



**Riya Sharma**-Mtech in E-Security and B.E in information technology from Bhilai Institute of Technology, Durg, Chhattisgarh. published two papers in International journal of management, technology and engineering Science.  
Email [riya2304sharma@gmail.com](mailto:riya2304sharma@gmail.com)



**Dr. (Mrs.) Ani Thomas** Professor and Head of the Department of Information Technology in Bhilai Institute of Technology, Durg. Chhattisgarh, She obtained her Masters degree in Computer Applications from Govt Engg. College, Raipur and PhD degree in 2013 from Csvtu . She has many reputed publications in her credit: 13 in international journals, 2 in national journals, 9 in International Conferences and 19 in National Conferences.

Email- [ani.thomas@bitdurg.ac.in](mailto:ani.thomas@bitdurg.ac.in)