International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

SALES PREDICTION IN TOURISM INDUSTRY USING DATA MINING

[1]Ms. Simaran Alam Mulani, [2]Ms. Rutuja Rajesh Malwadkar, [3]Ms. Alisha Mustak Pathan, [4]Mr. Pranav Ramdas Wagh, [5]Prof. Sayyad G.G

[1-4]Student, Department of Computer Engineering, SPVP's S.B. Patil College of Engineering, Indapur-413106

[5]Assistant Professor, Department of Computer Engineering, SPVP's S.B. Patil College of Engineering, Indapur-413106

Email id: [1]*simranmulani687@gmail.com*

Abstract:- *Using a significant quantity of data gathered from a variety of sources, the tourism and travel industry is working to improve the quality of its services. Because of how simple it is to acquire the opinions, assessments, and experiences of a variety of visitors, the planning of tourism has become more sophisticated and rich. Therefore, one of the biggest challenges that the tourism industry has is figuring out how to utilise the data that has been collected to identify the preferences of tourists. Unfortunately, some of the comments made by users are both irrelevant and difficult to interpret, making it difficult to make recommendations based on them. Aspect-based sentiment categorization algorithms have shown considerable potential in terms of overcoming the noise. There hasn't been a lot of study done on aspect-based emotion combined with categorization so far. The purpose of this study is to propose a framework for an aspect-based sentiment classification and recommendation system. Not only will it identify the aspects in a highly efficient manner, but it will also be able to complete classification tasks with a high level of accuracy utilising machine learning methods such as naive Bayes and Decision Tree. The performance of the framework has been tested by running tests using real-time datasets from Yelp and foursquare. The framework assists travellers in finding the best location, hotel, and restaurant in a city.*

Keyword*: Travel and tourism industry, Travel pattern, Classification, Clustering Data mining*

1. INTRODUCTION

The travel sector is one that is rapidly expanding, and it is becoming more important to some areas and countries as a primary industry. Each year, millions of tourists go to tourist attractions and share their impressions of those destinations on a variety of websites, including TripAdvisor and Opinion Table. These feelings provide an overall picture on the opinion bearer with relation to the tourism destination. However, there are a large number of speculations that are available about a certain location, and it is difficult for a regular user to examine/read all of these available evaluations and decide whether or not to visit a location. A variety of sentiment mining procedures have been offered as a means of managing the enormous number of hypotheses, and these methods assist in classifying the findings as either positive or negative. In any event, the newly presented methods do not take into account the many points of view that are associated with a sensation. Rather, the purpose of these tactics is to draw attention to the overarching ideas shared by all points of view. Following that, new strategies for opinion mining based on elements of the opinions were presented. With the use of these procedures, clients are able to disentangle their varied perspectives from their emotions and classify each perspective included in the evaluations as either good or negative. Take, as an example, the following sentence: "Nourishment is delightful, yet administration is slow[11]. "The terms "nourishment" and "administration" make reference to points of view, with "flavorful" being a favourable evaluation of the "nourishment" perspective and "moderate" representing a less favourable evaluation of the "administration" viewpoint.

When it comes to extracting aspects, one of the first challenges is to determine which aspects are implicit. Implicit features are those that don't make a direct appearance in any viewpoint, yet they nonetheless point to a crucial component. For example, in the provided phrase, "yesterday my sister and I visited Sayaji Hotel, the taste was superb," the user did not describe any vital component in this sentence[12]. This is an example of a given statement. However, the "food" part is strongly hinted to by the implication of this line. The second challenge is the difficulty in determining which elements are referential. It is very uncommon for individuals to utilise a variety of terms and phrases while attempting to communicate the same concept. For the purpose of describing the setting of a restaurant, for instance, the terms "atmosphere" and "ambiance" relate to the same thing, and the two terms are interchangeable[13]. Thirdly, determining the infrequent characteristics may be a very difficult and time-consuming process. Because there was such a huge number of explicit aspects, the existing technique for aspect extraction simply ignored the less

common ones. yet, certain uncommon features may be co-referential of frequent aspects or may be vital for a tourist destination; for instance, air conditioner and bed are examples of less frequent aspects; yet, these aspects are important for hotels[14].

Through the presentation of innovative machine learning techniques, this study demonstrates a robust framework for aspect-based estimate order. The structure is made up of two fundamental components: a choice tree-based viewpoint recognisable proof strategy that enables readers to recognise explicit, implicit, and infrequent aspects, and groups of co-referential aspects from tourist sentiments that are classified using aspect-based sentiment classification using machine learning algorithms that have three stages[15]. The Stanford Basic Dependency technique is used in the main stage to channel sentence components between slant words and aspects in a particular opinion statement. This is done in order to better express the argument. During the second step of the process, filtered phrases are used to construct features such as n-grams and Part-of-Speech tags. In the last step, methods for machine learning are used to find attributes for the purpose of categorising the views on elements as either good or bad[16].

2. LITERATURE SURVEY

The field of information mining has seen a great deal of study during the course of its existence. A piece of writing that was assessed discusses the advantages of utilising community information in projecting demand for trips to China. Mapping information business enterprise and network of relationships to build demand in the tourism region among the author, and describes using strategies of information mining web, analyses and forecasting of tourism demand, primarily based on the precept of building expertise and motives internet information mining procedure evaluation and forecasting. In the event of doing an empirical study utilising Shanghai as the metropolis, the following things should be considered.

People in the community have the possibility to find work thanks to travel and tourism. The sector of travel and tourism in India is expanding at a breakneck speed[17]. Employment for an uncountable number of individuals. Many popular tourist locations are visited by a respectable number of people from other countries. This contributes to the country's earning of exchange. There has been a substantial quantity of writing produced in the fields of travel and tourism and the elements that contribute to it as a result of the many research that have been carried out in different parts of the world. A significant number of them are

backed by various features. Greater economic advantages, including employment, revenue, and the production of foreign generations, are provided by the tourism industry within the area[17]. No pertinent information is provided to the traveller or tourist[24].

This collection of rules use iterative methods to group examples from a dataset into clusters with features that are comparable to one another. These classifications are helpful for investigating the data, locating irregularities within the records, and formulating forecasts. This first determines the connections that exist within a dataset, and then it creates a series of clusters that are mostly based on the linkages that have been found[23]. As shown in the accompanying picture, a scatter plot is an effective method for graphically representing how an algorithm generates statistics. This method has many applications. Every single one of the instances included in the dataset is shown as a factor in the scatter plot, which depicts all of the cases. The clusters organise the components shown on the graph into groups, and they indicate the connections that the set of rules finds between them[18].

A class and regression procedure may be derived from the decision timber set of rules. The predictions that the algorithm produces for discrete qualities are mostly determined by the relationships that exist between the input columns in a dataset. It makes use of the values of those columns, which are referred to as states[22], to anticipate the states of a column that you have designated as predictable. In specifically, the set of rules determines the columns that may be entered that have the potential to be connected with the column that is predicted. A histogram may be used to verify the manner in which a collection of rules constructs a tree for a discrete predictable column. This validation can take place[21]. The following figure illustrates an example of a histogram that compares a column that can be predicted, bike shoppers, to a column that can be input, age. The histogram demonstrates that the age of a person makes it possible to determine whether or not that man or woman will purchase a bicycle[19].

Sequential sample Mining reveals exciting sequential patterns in some of the enormous datasets. It unearths out frequent subsequence's as patterns from a sequence database. Mining sequential patterns from a database is becoming more interesting to many different types of businesses as a result of the vast amounts of data that are continually being collected and stored. Sequential pattern mining is one of the most well-known approaches, and it has a wide range of applications. Some examples of these applications include

internet-log analysis, assessment of consumer buying behaviour, and evaluation of medical documents. The transactional information of customers may be mined for sequential patterns in the retail industry. This enables the business to better serve its customers[20]. As an example, a client who previously purchased a notebook returns to make a second purchase of a PDA in addition to a WLAN card. The business is able to read the behaviour of the customers, understand their interests, meet their wants, and precisely anticipate their desires with the use of such data as they may be used to read the customers' behaviour.

3. PROPOSED SYSTEM

Data mining methods have been deemed the greatest method for predicting sales in the tourist business, despite the fact that there are a number of different forecasting models that may be used to determine sales in the tourism sector. Data mining is the process of discovering valuable patterns, correlations, and rules that were not previously known by sifting through a vast quantity of data that is kept in some repository (database). This process is characterised as the "finding out useful patterns, correlations, and rules that were not previously known." The forecast of sales in the tourist business is handled by this system, which takes into account two different data sets. The first data set is a count of orders, while the second data set is sales.
2. An evaluation of the remarks' overall tone

At this point, the orders table is analysed in order to determine the frequency with which a certain package has been selected by the user. After that, the orders database is used to do the counting necessary to determine the preferences of the users.

The second kind of analysis is called sentiment analysis, and it involves scanning the words that users write into their comments in order to determine if those words are favourable or negative. In accordance with this, the scores are determined. After the ratings have been given out, the scores are summed together to determine the overall rating. Consequently, the orders table and the comments table are the two data sets that are utilised in this instance in the process of forecasting sales in the tourist business.

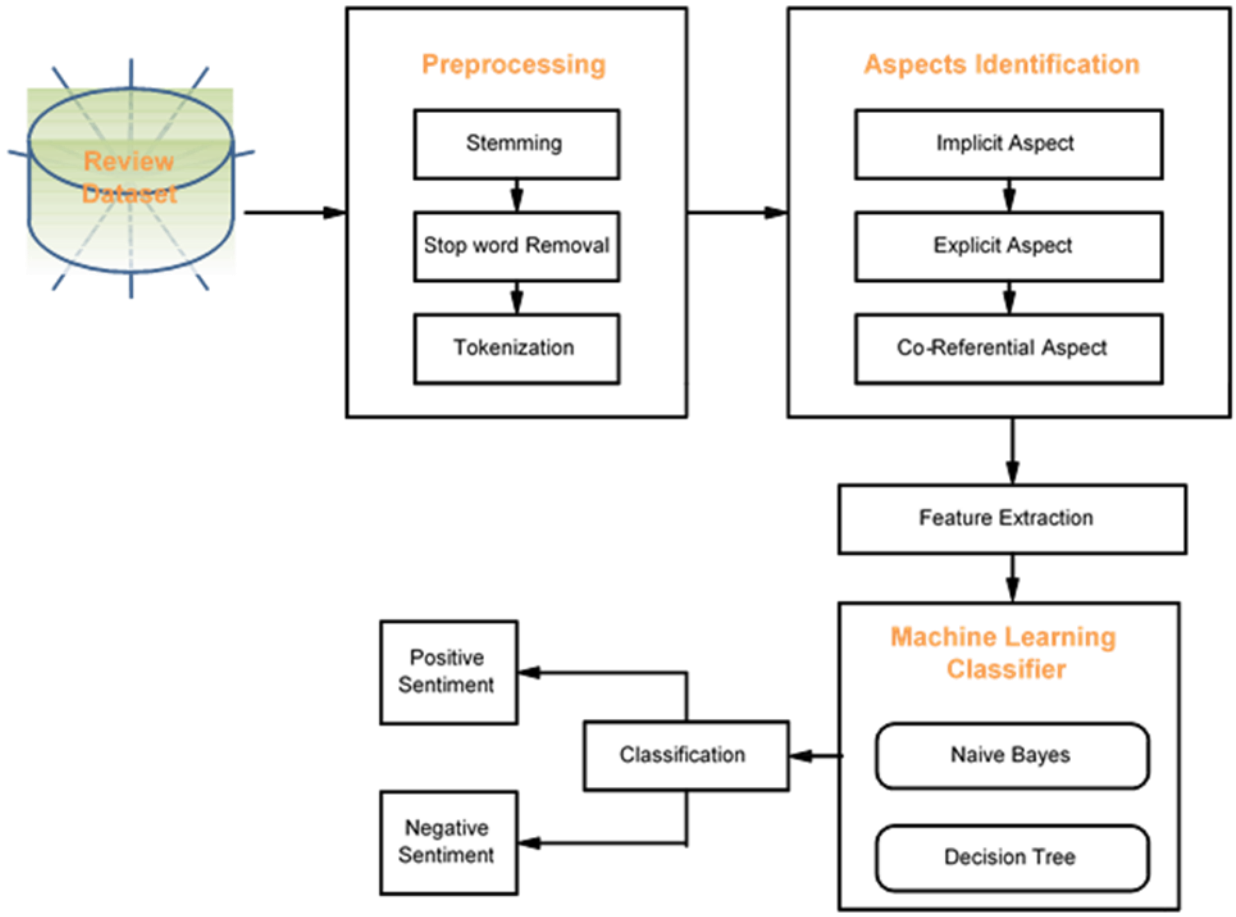

Fig 1: System Architecture

Proposed System is divided in to 2 parts

Admin

Add Packages- Here the admin will add place details along with other package information.

- Add reviews-Admin can add reviews
- View Reviews- Admin can view the reviews
- Sales report- Graphical representation of the sales report is made available to the user

User

- View Packages-Here the user can view the packages added by the admin and can book the same
- Send Feedback – here the user can add feedback

Applications:

- Recommendation Applications
- Tourist Applications
- Sales Prediction applications

4. CONCLUSION

A framework for aspect-based sentiment categorization was described in this system's proposal. This framework sorts reviews of aspects into either positive or negative categories. A technique for the extraction of aspects using a tree-based structure is presented inside this framework. This approach can extract both explicit and implicit aspects from tourist opinions. It will first extract common nouns and noun phrases from the reviews text, and then it will use Word-Net to group nouns that are related. When conducting reviews, the decision tree method is used, in which the review words are utilised as internal nodes and the extracted noun is used as the leaf of the tree. In the initial step of the process, the Stanford Basic Dependency analysis is performed on each phrase to weed out opinion less and irrelevant statements. The last step is to extract features from the remaining sentences using N-Grams and POS Tags in order to train the classifiers. Last but not least, in order to train the classifiers, machine learning techniques are applied to the characteristics that were retrieved.

Reference

1. Shuyao Qi, Dingming Wu, and Nikos Mamoulis," Location Aware Keyword Query Suggestion Based on Document Proximity" VOL. 28, NO. 1, JANUARY 2016
2. X. Liu, Y. Liu, and X. Li, "Exploring the context of locations for personalized Location recommendations," in Proceedings of IJCAI'16. AAAI, 2016.
3. H. Li, R. Hong, D. Lian, Z. Wu, M. Wang, and Y. Ge, "A relaxed ranking-based factor model for recommender system from implicit feedback," in Proceedings of IJCAI'16, 2016, pp. 1683–1689.
4. D. Lian, Y. Ge, N. J. Yuan, X. Xie, and H. Xiong, "Sparse Bayesian collaborative filtering for implicit feedback," in Proceedings of IJCAI'16. AAAI, 2016.
5. X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," in Proceedings of SIGIR'16, vol. 16, 2016.

6. F. Yuan, G. Guo, J. M. Jose, L. Chen, H. Yu, and W. Zhang, "Lambdafm: learning optimal ranking with factorization machines using lambda surrogates," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016, pp. 227–236.
7. Yiding Liu1 Tuan Anh Nguyen Pham2 Gao Cong3 Quan Yuan," An Experimental Evaluation of Point of interest Recommendation in Location based Social Networks-2017"
8. Shuhui Jiang, Xueming Qian *, Member, IEEE, Tao Mei, Senior Member, IEEE and Yun Fu, Senior Member, IEEE" Personalized Travel Sequence Recommendation
9. on Multi-Source Big Social Media" Transactions on Big Data IEEE TRANSACTIONS ON BIG DATA, VOL. X, NO. X,
10. Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo "Personalized Travel PackageWith Multi-Point-of-Interest Recommendation Based on Crowdsourced User Footprints" 2016
11. Salman Salamatian_, Amy Zhangy, Flavio du Pin Calmon_, Sandilya Bhamidipatiz, Nadia Fawazz, Branislav Kvetonx, Pedro Oliveira{, Nina Taftk "Managing your Private and Public Data: Bringing down Inference Attacks against your Privacy" 2015
12. AA Khan, RM Mulajkar, VN Khan, SK Sonkar, DG Takale. (2022). A Research on Efficient Spam Detection Technique for IOT Devices Using Machine Learning. NeuroQuantology, 20(18), 625-631.
13. SU Kadam, VM Dhede, VN Khan, A Raj, DG Takale. (2022). Machine Learning Methode for Automatic Potato Disease Detection. NeuroQuantology, 20(16), 2102-2106.
14. DG Takale, Shubhangi D. Gunjal, VN Khan, Atul Raj, Satish N. Gujar. (2022). Road Accident Prediction Model Using Data Mining Techniques. *NeuroQuantology*, *20*(16), 2904-2101.
15. SS Bere, GP Shukla, VN Khan, AM Shah, DG Takale. (2022). Analysis Of Students Performance Prediction in Online Courses Using Machine Learning Algorithms. NeuroQuantology, 20(12), 13-19.
16. R Raut, Y Borole, S Patil, VN Khan, DG Takale. (2022). Skin Disease Classification Using Machine Learning Algorithms. NeuroQuantology, 20(10), 9624-9629.
17. SU Kadam, A katri, VN Khan, A Singh, DG Takale, DS. Galhe (2022). Improve The Performance Of Non-Intrusive Speech Quality Assessment Using Machine Learning Algorithms. NeuroQuantology, 20(19), 3243-3250.
18. DG Takale, (2019). A Review on Implementing Energy Efficient clustering protocol for Wireless sensor Network. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 310-315.

19. DG Takale. (2019). A Review on QoS Aware Routing Protocols for Wireless Sensor Networks. International Journal of Emerging Technologies and Innovative Research, Volume 6(Issue 1), 316-320.
20. DG Takale (2019). A Review on Wireless Sensor Network: its Applications and challenges. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 222-226.
21. DG Takale, et. al (May 2019). Load Balancing Energy Efficient Protocol for Wireless Sensor Network. International Journal of Research and Analytical Reviews (IJRAR), 153-158.
22. DG Takale et.al (2014). A Study of Fault Management Algorithm and Recover the Faulty Node Using the FNR Algorithms for Wireless Sensor Network. International Journal of Engineering Research and General Science, Volume 2(Issue 6), 590-595.
23. DG Takale, (2019). A Review on Data Centric Routing for Wireless sensor Network. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 304-309.
24. DG Takale, VN Khan (2023). Machine Learning Techniques for Routing in Wireless Sensor Network, IJRAR (2023), Volume 10, Issue 1.