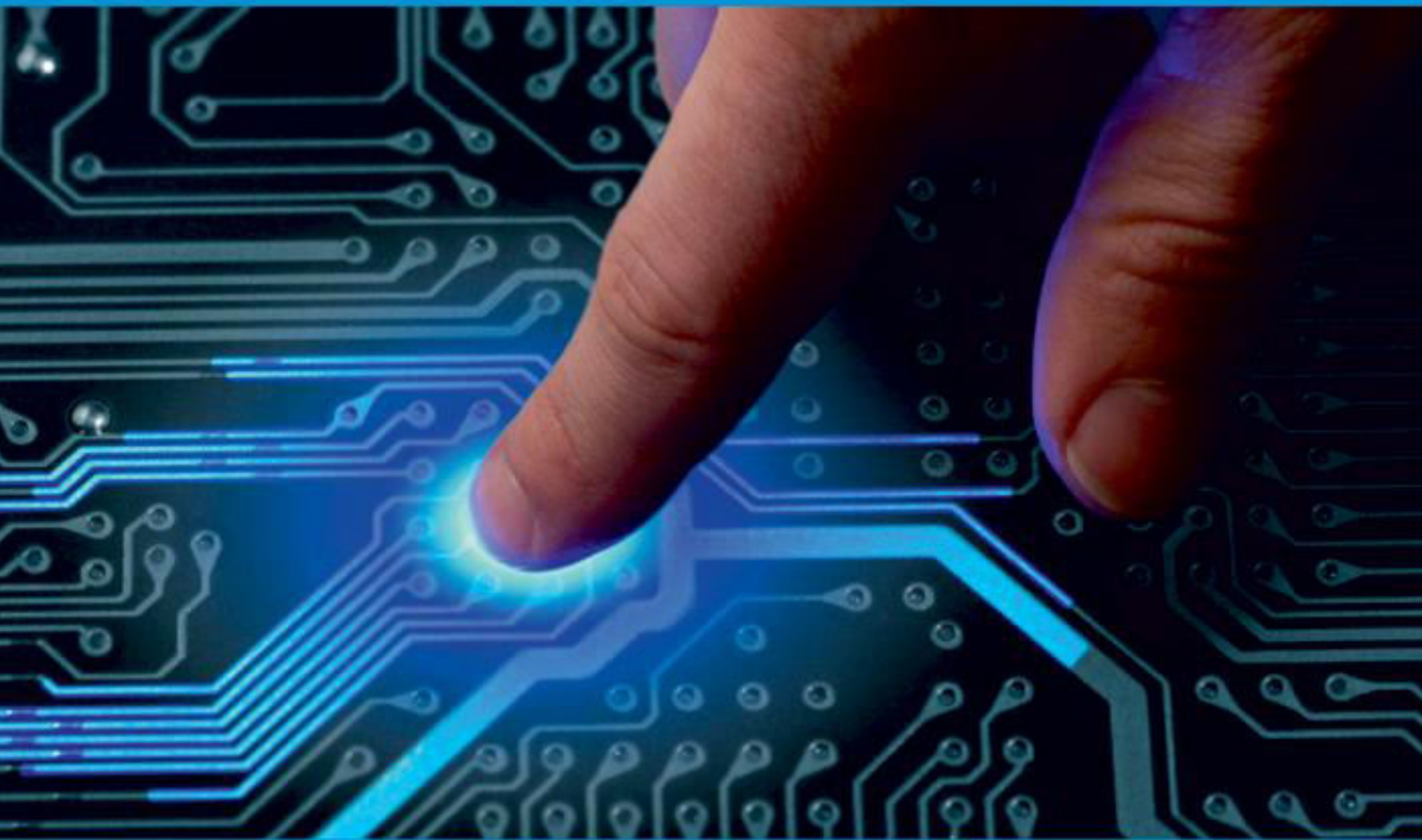
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

Sales Prediction in Tourism Industry using Machine Learning

Pragati Malusare, Arun Nishad, Karan Shinde, Ankit Kumar, Vivekanand Singh

Department of Computer Engineering, Indira College of Engineering and Management, Pune, India

ABSTRACT*:* The tourism is using a vast amount of data obtained from various sources to improve services. The simple availability of feedback, evaluations, and impressions from various visitors has made tourism planning rich and complex. As a result, using the gathered data to detect tourist tastes is a major challenge for the tourism industry. Unfortunately, some user comments are meaningless and difficult to understand, making recommendations difficult. Sentiment classification approaches based on aspects have shown promise in eliminating noise. There isn't a lot of work on aspect-based sentiment with classification right now. This paper introduces a framework for an aspect-based sentiment classification recommendation method that uses deep learning algorithms to not only classify aspects quickly, but also to perform classification tasks with high accuracy. The framework assists tourists in locating the best location, hotel, and restaurant in a region, and its effectiveness has been assessed through experiments on real-time review classification.

KEYWORDS: Deep Learning, Classification, Reviews, Aspect Based Sentiment Analysis, Text analysis.

I. INTRODUCTION

The tourism is a rapidly growing industry that is important to some regions and countries as a main industry. Every year, millions of people visit tourist attractions and share their experiences on websites like Trip Advisor and Opinion Table. These sentiments provide a general viewpoint on a person's feelings about a tourist attraction. In either case, there are several suppositions available on a particular location, and it is difficult for a typical consumer to audit/read all of these available evaluations and decide whether or not to visit a location. Completely different perspective analysis methods have been proposed to deal with the large number of hypotheses, and these techniques aid in categorizing the results as positive or negative. In either case, the approaches that have recently been suggested do not deal with the various viewpoints that exist in a feeling. Rather, these tactics actually draw attention to the common feelings of all points of view. Following that, current opinion processing methods depend on aspects were introduced. Clients may use these techniques to separate different perspectives from emotions and classify each perspective in the evaluations into positive and negative categories. "Nourishment is delightful, but administration is sluggish," for example, in a simple word. First and foremost, defining the implicit aspects is a challenge in terms of aspects extraction. Implicit aspects do not appear clearly in any view, but they do suggest a significant aspect. For example, the user did not mention any important aspects in the sentence **"yesterday my sister and I visited mahendra hotel, the taste was superb."** However, the implication of this sentence is that it has something to do with food. Second, determining the coreferential elements is challenging. People frequently use a variety of terms and phrases to explain the same thing. In a restaurant, for example, atmosphere and ambience apply to the same thing, and they are mutually exclusive. Third, recognizing the uncommon features is time-consuming. The infrequent aspects were discarded by usable aspect extraction methods due to the large number of explicit aspects. Some infrequent aspects, on the other hand, can be coreferential to frequent aspects or relevant for a tourist destination; for example, air conditioning and beds are less frequent aspects, but they are important for hotels.

By introducing advanced deep learning algorithms, this paper demonstrates a powerful mechanism for aspect-based estimation order. There are two basic components to the structure. option tree-based perspective recognizable proof strategy, which helps readers to recognize formal, implicit, and infrequent aspects, as well as groups coreferential aspects from tourist sentiments aspect-based emotion classification using deep learning algorithms with three stages The Stanford Basic Dependency technique is used in the main stage to channel sentence sections in a given opinion sentence between slant words and aspects. Filtered phrases are used in the second stage to create features like n-grams and Part-Of-Speech tags. Finally, deep learning algorithms are used to recognize characteristics that can be used to categorize positive and negative opinions about various aspects.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

II. LITERATURE REVIEW

M. Colhon et.al [1] presented sentiment classification system for categorizing tourist reviews based on the sentiment. Authors also presented the findings of a real-world implementation of the proposed sentiment analysis process. This information is taken from the AmFostAcolo tourist review website. It focuses on determining the relationship between the holder of an opinion and the quality of that opinion, the sentiment of the review with the review score. As a consequence of authors observations, it is concluded that some attributes of the opinion holder, such as His or her integrity, for example, may be linked to the accuracy of the views shared in his or her evaluations.

A. Mukherjee et.al [2] presented different environment in this paper, where the consumer provides some seed words for a few aspect categories and the model extracts and clusters aspect terms into categories at the same time. This setting is critical since categorizing aspects is a subjective process that may require different categorizations depending on the application. It is desirable to provide some kind of user guidance. The authors of this paper suggest two mathematical models to solve this seeded problem, with the aim of determining exactly what the consumer desires. Authors proposed two models SAS and MESAS which take seeds reflecting the user needs to discover specific aspects. ME-SAS also does not need any additional help from the user in its Max-Ent training. Our results showed that both models outperformed two state-of-the-art existing models ME-LDA and DF-LDA by large margins.

L. Zhang and B [3]. Liu presented the computational study of people's views, appraisals, perceptions, and emotions toward entities such as goods, programmes, organizations, persons, events, and their various aspects is known as opinion mining or sentiment analysis. Natural language analysis and Web mining have become active research areas.in the past two years Opinion mining has been researched by researchers at the document. Levels of sentence and Aspect-level opinion mining (also known as aspect-based opinion mining) is often needed in practical applications because it offers comprehensive opinions or facts. Sentiments regarding multiple aspects of entities, as well as entities themselves normally, intervention is taken. As a result, there are two types of extraction: aspect extraction and object extraction.

L. Rosa et.al [4] proposed a music recommendation system based on an evaluation force metric called improved Sentiment Metric (eSM), which is the relationship between a vocabulary-based estimation metric and a client-specific remedy factor. Methods for abstract experiments, led in a research centre condition, are used to discover this remedy factor. The remedy factor is specified and used to change the last supposition force based on the test results. The music proposal process is conducted through a method of low multifaceted existence for mobile phones, and the clients' assumptions are isolated from sentences posted on interpersonal organizations, which suggests melodies based on the slant force of the current client similarly, the structure was built with ergonomics and ease of use in mind.

R. Moraes et.al [5]: An empirical comparison of SVM and ANN for document-level sentiment analysis is presented. Authors addressed the criteria, models that result, and situations in which both methods improve classification accuracy. In a typical bag-of-words model, they used a standard evaluation context and common supervised methods for feature selection and weighting. Their experiments showed that, with the exception of a few unbalanced data contexts, ANN produces superior or at least equivalent results to SVMs. Even in the light of unbalanced results, ANN outperformed SVM by a statistically significant difference on the benchmark dataset of Movies reviews.

G. Wang et.al [6]: In this authors proposed that User-generated content can be quickly shared publicly due to the exponential progress in information technology. Although individuals, companies, and governments are interested in assessing the sentiments behind this content, no consensus exists on which sentiment classification technologies are the most successful. Recent research Ensemble learning approaches may have potential use in emotion classification, according to the researchers. Author compared the performance of three commonly used ensemble methods (Bagging, Boosting, and Random Subspace) focused on five different base learners (Naive Bayes, Maximum Entropy, Decision Tree, and K).For sentiment classification, Nearest Neighbor and Help Vector Machine are used. Furthermore, ten public opinion surveys were conducted.

E. Marrese-Taylor et.al [7]: Authors presented an extension of Bing Liu's viewpoint-based feeling mining method for use in the travel industry. The extension is concerned with how customers refer to various types of products in different ways when filling out online surveys. Since Liu's approach is based on physical item audits, it couldn't be extended directly to the travel industry, which has features that aren't taken into account by the model. Arrangement at the viewpoint stage. These highlights were discovered through an itemized investigation of on-line travel industry item surveys, and authors then modelled them in their expansion, proposing the use of new and increasingly complex NLP-

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

based criteria for abstract and supposition arrangement at the viewpoint stage. Involve the project of feeling awareness and list, as well as suggest new techniques to assist clients in processing the enormous accessibility of feelings in a straightforward manner.

Z. Hai et.al [8]: The authors of this paper proposed a novel method for identifying opinion features from online reviews by leveraging the disparity in opinion feature statistics between two corpora, one domain-specific (i.e., the provided review corpus) and one domain-independent corpus (i.e., the contrasting corpus). This difference is captured using a metric called domain relevance (DR), which characterizes a term's relevance to a text set. By specifying a set of syntactic dependency laws, authors first extracted a list of candidate opinion features from the domain analysis corpus. On the domain-dependent and domain-independent corpora, they estimated intrinsic-domain relevance (IDR) and extrinsic-
domain relevance (EDR) scores for each extracted candidate function. Opinion characteristics are then validated if they are less generic (EDR score less than a threshold) and more domain-specific (IDR score greater than another threshold).

C. S. Khoo and S. B. Johnkhan [9]: Authors proposed a new general-purpose sentiment lexicon called WKWSCI Sentiment Lexicon and compares it with five existing lexicons: Hu Liu Opinion Lexicon, Multi-perspective Question Answering (MPQA) Subjectivity Lexicon, Word-Sentiment Association Lexicon and Semantic Orientation Calculator (SO-CAL) lexicon from the National Research Council of Canada (NRC). Using an Amazon product review data set and a news headlines data set, the usefulness of the emotion lexicons for sentiment categorization at the text and sentence stage was analyzed. When acceptable weights are used for various types of sentiment terms, WKWSCI, MPQA, Hu Liu, and SO-CAL lexicons are similarly good for product review sentiment categorization, with accuracy rates of 75%–77%. When a teaching corpus isn't available, Hu Liu found that counting positive and negative terms for both document-level and sentence-level sentiment categorization yielded the best results.

M. Afzaal et.al [10]: Authors proposed a fuzzy aspect-based opinion classification scheme that derives aspects from user views effectively and performed classification that is close to correct. To assess the feasibility of their suggested scheme, they ran tests on real-world datasets. The suggested scheme not only extracts aspects well, but also increases classification accuracy, according to the findings.

III. PROPOSED SYSTEM ARCHITECTURE

The proposed framework for aspect recognition and classification is presented in architecture.

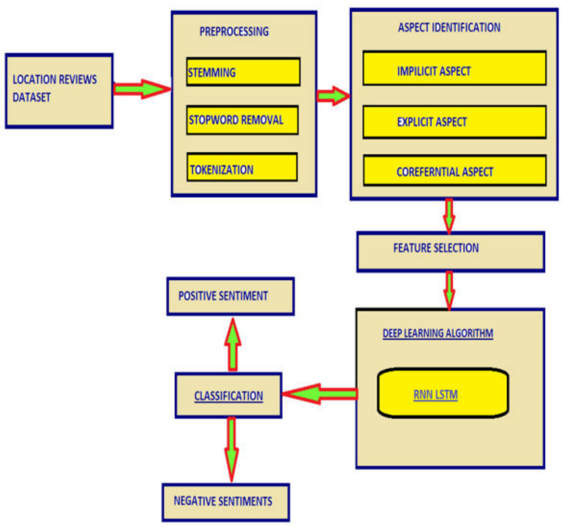


Fig. 1. Proposed System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

A deep learning algorithm classifies each aspect in a consumer review into positive or negative by considering all aspects and their linkages to sentiment words. For example, in a restaurant review, the tourist likes the food but dislikes the service. The class of this review depends on the sentiment words and phrases linked to aspects. When multiple aspects are considered, the situation becomes more complex; deep learning algorithms are very efficient and helpful.

D. Algorithm

Recurrent Neural Network (LSTM):-

RNNs are a form of Neural Network in which the output from the previous step is used as input in the current step. All of the inputs and outputs of standard neural networks are independent of one another, but in certain situations, such as when predicting the next word of a sentence, the previous words are necessary, and therefore the previous words must be remembered. As a result, RNN was developed, which used a Secret Layer to solve the problem. The Secret condition, which remembers any details about a sequence, is the most important function of RNN.

IV. RESULT AND ANALYSIS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i5-2120 CPU @ 3.30GHz, 8GB memory, Windows 10, MySQL backend database and jdk 1.9. The application is dynamic web application for design code in Eclipse IDE and execute on Tomcat server 8.0.

The overall accuracy of LSTM Recurrent Neural Network classification technique is performed. So this works gives better

Classification results.

Calculation Formula:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance), TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

On the basis of this parameter, we can calculate four measurements

Accuracy = TP+TN/TP+FP+TN+FN Precision = TP /TP+FP Recall= TP/TP+FN

The data analysis for the performance of LSTM RNN is –

Total samples = 155

Here it is found -

True Positive=90
False Positive=10
True Negative=50
False Negative=5

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

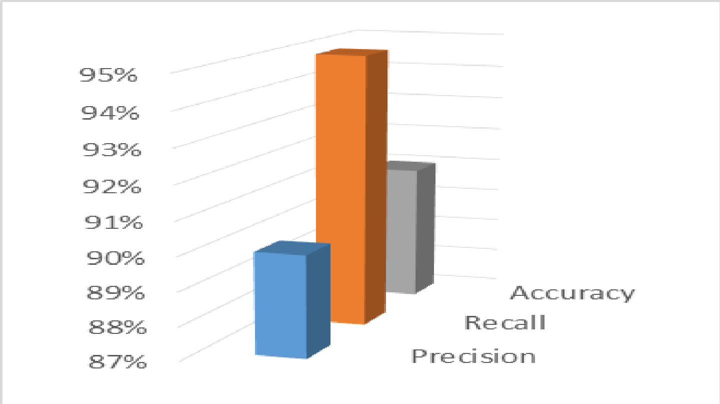


Fig-10- Performance Analysis Graph of LSTM RNN

	Recurrent Neural Network
Precision	90%
Recall	95%
Accuracy	91%

V. CONCLUSION

This paper introduced an aspect-based sentiment classification system for categorizing positive and negative reviews about aspects. A tree-based aspects extraction approach that removes both overt and tacit aspects from tourist opinions is suggested in this context. It removes common nouns and noun phrases from the text of reviews. On feedback, a LSTM recurrent neural network is used. Stanford Basic Dependency is used on each sentence to delete opinion-free and meaningless sentences. To train the classifiers, features are extracted from the remaining sentences using N-Grams and POS Tags. Finally, deep learning algorithms LSTM are used to train the classifiers using the derived features. The learned model is used to differentiate the sentiment regarding derived aspects into positive or negative.

REFERENCES

1. M. Colhon, C. Badicˇa,ˇ and A. S¸endre, ”Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews,” in Int. Conf. Knowledge Sci., Eng. and Manage., 2014, pp. 246-257, DOI: 10.1007/978-3-319-12096-622
2. A.Mukherjee and B. Liu, ”Aspect extraction through semisupervised modeling,” in Proc. 50th Annu. Meeting Assoc. for Computational Linguistics, 2012, pp. 339-348.
3. L. Zhang and B. Liu, “Aspect and entity extraction for opinion mining,” in Data Mining Knowledge Discovery For Big Data, Berlin, Germany: Springer, 2014, pp. 1-40, DOI: 10.1007/978-3- 642-40837-3
4. R. L. Rosa, D. Z. Rodriguez, and G. Bressan, “Music recommendation system based on user’s sentiments extracted from social networks,” IEEE Trans. Consum. Electron., vol. 61, no. 3, pp. 359-367, Aug. 2015, DOI: 10.1109/TCE.2015.7298296
5. R. Moraes, J. F. Valiati, and W. P. G. Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Ex-pert Sys. With Appli., vol. 40, no. 2, pp. 621-633, Feb. 2013, DOI: 10.1016/j.eswa.2012.07.059
6. G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, “Sentiment classification: The contribution of ensemble learning,” Decision Supp. Sys., vol. 57, pp. 77-93, Jan. 2014, DOI: 10.1016/j.dss.2013.08.002
7. E. Marrese-Taylor, J. D. Velasquez,´ and F. Bravo-Marquez, “A novel de-terministic approach for aspect-based opinion mining in tourism products reviews,” Expert Sys. With Appli., vol. 41, no. 17, pp. 7764-7775, Dec. 2014, DOI: 10.1016/j.eswa.2014.05.045
8. Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, “Identifying features in opinion mining via intrinsic and extrinsic domain relevance,” IEEE Trans. Know. And Data Engi., vol. 26, no. 3, pp. 623-634, Mar. 2014, DOI: 10.1109/TKDE.2013.26

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105186 |

9. C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," Jour. Inform. Scien., vol. 44, no. 4, pp. 491-511, Aug. 2018, DOI: 10.1177/0165551517703514
10. M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, "Fuzzy Aspect Based Opinion Classification System for Mining Tourist Reviews," Adva. In Fuzzy Sys., vol. 2016, Oct. 2016, DOI:10.1155/2016/6965725