iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40865

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Recommendation System for Tourist Reviews using Aspect Based Sentiment Classification

Kande Trupti V[1], H. P. Shah[2]
[1, 2]*Computer Science and Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere*

Abstract: ***To improve services, the tourism industry makes use of a large amount of data collected from a variety of sources. Because of the easy availability of feedback, evaluations, and impressions from a wide range of visitors, tourism planning has become both rich and complex. As a result, the tourism industry faces a significant challenge in determining tourist preferences based on the data collected. Unfortunately, some user comments are meaningless and difficult to comprehend, making it difficult to make recommendations. Approaches to sentiment classification that are based on aspects have shown promise in terms of reducing noise. At the moment, there isn't a lot of work being done on aspect-based sentiment and classification. Aspect-based sentiment classification recommendation methods are introduced in this paper, which employ deep learning algorithms to not only classify aspects quickly, but also to perform classification tasks with high accuracy. A series of experiments on real-time review classification have been conducted to determine how effective the framework is at assisting tourists in locating the best location, hotel, and restaurant in a region.***

Keywords: Classification, Deep Learning, Tourist Reviews, Aspect Based Sentiment Analysis.

I. INTRODUCTION

Tourism is a rapidly expanding industry that is important as a primary industry in some regions and countries. Every year, millions of people travel to tourist destinations and share their impressions of their adventures on websites such as Trip Advisor and Opinion Table. When combined, these sentiments provide a general perspective on a person's feelings toward a tourist attraction. There are several hypotheses available on a particular location in either case, and it is difficult for the average consumer to audit/read all of the available evaluations and determine whether or not to visit a location in either case. To deal with the large number of hypotheses, completely different perspective analysis methods have been proposed, and these techniques aid in categorizing the results as positive or negative. In either case, the approaches that have recently been proposed do not deal with the various points of view that exist in a sense of what is happening. Rather, these tactics serve to draw attention to the feelings that are shared by people of all political persuasions. Following that, the current opinion processing methods that are based on aspects were presented to the audience. Using these techniques, clients may be able to distinguish between different perspectives and emotions, as well as categorize each perspective in the evaluations into positive and negative categories. "Nourishment is wonderful, but the administration is sluggish," to put it another way, is a simple statement. First and foremost, defining the implicit aspects is a difficult task in terms of aspects extraction because they are not explicitly stated. Implicit aspects do not appear clearly in any view, but they do suggest a significant aspect that should be taken into consideration. For example, in the sentence "yesterday my sister and I visited Mahindra hotel, the taste was superb," the user did not mention any significant details about the experience. The implication of this sentence, on the other hand, is that it has something to do with eating. In addition, determining the coreferential elements is a difficult task. Many different terms and phrases are frequently used to explain the same thing by different people. Atmosphere and ambiance are synonymous in a restaurant, but they are not the same thing, and they are not mutually exclusive either. Third, it takes time to recognize the features that are unique to the situation. Because of the large number of explicit aspects, usable aspect extraction methods were unable to identify the infrequent aspects and discarded them. 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. Air conditioning and beds are not the only infrequent aspects that can be coreferential to frequent aspects. This paper, which introduces advanced deep learning algorithms, demonstrates a powerful mechanism for aspect-based estimation order by introducing advanced deep learning algorithms. The structure is comprised of two fundamental components. aspect-based emotion classification using deep learning algorithms with three stages: In the main stage, the Stanford Basic Dependency technique is used to channel sentence sections in a given opinion sentence between slant words and aspects using the Stanford Basic Dependency technique. Second-stage features such as n-grams and Part-of-Speech tags are created by using filtered phrases in the first stage. Last but not least, deep learning algorithms are used to recognize characteristics that can be used to categorize positive and negative opinions about a variety of topics.

©IJRASET: All Rights are Reserved

II. RELATED WORK

M. Colhon et.al [1] presented sentiment classification system for categorizing tourist reviews based on the sentiment. The 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 the author's 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 a 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 solve this seeded problem, with the aim of determining exactly what the consumer desires. The 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, programs, organizations, persons, events, and their various aspects are 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) are 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 center 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. The 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.

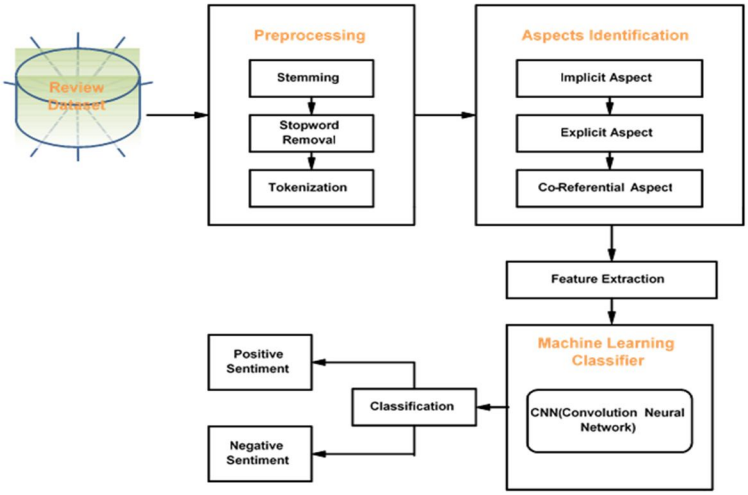
G. Wang et.al [6]: In this authors proposed that User-generated content can be quickly shared publicly due to the exponential progress in information technology. Although individuals, companies, and governments are interested in assessing the sentiments behind this content, no consensus exists on which sentiment classification technologies are the most successful. Recent research Ensemble learning approaches may have potential use in emotion classification, according to the researchers. The 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. The arrangement at the viewpoint stage. These highlights were discovered through an itemized investigation of online travel industry item surveys, and authors then modeled them in their expansion, proposing the use of new and increasingly complex NLP-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 compare 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.



1) Data Collection
2) Data Pre-processing
3) Aspect Identification
4) Classification

A. Review Data Collection:

In data collection, reviews are collected from popular social media websites using crawlers and APIs. The datasets have different numbers of reviews in each domain. In the restaurant domain, there are 2000 reviews with 1000 positive and 1000 negative. In the hotel domain, there are 4000 reviews with 2000 positive and 2000 negative. London is chosen as a city of interest in the case study.

B. Data Preprocessing

Data preprocessing removes redundancy and ambiguity inherent in the data and transforms the reviews into sentences to facilitate sentence-level aspect-based classification. First, sentences are extracted by identifying the delimiters (e.g. dot, exclamation, or question mark). Next, redundant information, e.g. duplicate sentences, is removed. Finally, ambiguous, vague, or misspelled terms are corrected.

C. Aspect Identification

The objective of the aspect identification method is to identify aspects that are important and relevant to a tourist place. This paper proposes a hybrid aspect identification method that can identify both explicit and implicit aspects from reviews about tourist places based on categorization.

D. Data Classification

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.

IV. ALGORITHM

A. Hybrid Tree-Based Aspect Identification

Input: Collection of sentences = {S1, S2, S3...Sn}
Output: Aspects assigned to sentences

1. **initialize** aspects
2. **for** all sentences **do**
3. Stanford tagger = SPOS (sentences)
4. **if** NN is in Stanford tagger **then**
5. Aspects ← NN
6. **end if**
7. **end for**
8. **initialize** aspects groups
9. **for** all aspects **do**
10. WordNet _sets = WNSS (aspects)
11. **if** TRUE in WordNet sets **then**
12. aspects groups ← aspects
13. **end if**
14. **end for**
15. frequent_aspects = freq _measure (aspects, aspects groups,10)
16. tree = DT (sentences, frequent aspects)
17. **initialize** aspect_assigned sentences
18. **for** all sentences **do**
19. aspect _identification = tree (sentences)
20. **if** TRUE in aspect identification then
21. aspect _assigned _sentences ← aspect identification
22. **end if**
23. **end for**
24. **return** aspect_assigned_sentences

B. Convolutional Neural Network

Steps:-

1. Consider a network of one input layer, three hidden layers, and one output layer. Each hidden layer, like other neural networks, would have its own collection of weights and biases, such as (w1, b1) for hidden layer 1, (w2, b2) for the second hidden layer, and

(w3, b3) for the third hidden layer. This suggests that each of these layers is self-contained, meaning that they do not remember the previous outputs.
2. The network is given a single time phase of the input. Then, using the current input and the previous state, Determine the current state.
3. For the next time stage, the current ht becomes ht-1.
4. According to the problem, one can go through as many time phases as required and merge the data from all previous states.
5. The final current state is used to measure the output after all of the time phases have been completed.
6. The error is then determined by adding the output to the real output, i.e. the target output.
7. The error is then passed back to the network, which updates the weights, and the network is therefore qualified.

V. RESULT AND DISCUSSION

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 a dynamic web application for design code in Eclipse IDE and executes on Tomcat server 8.0.
The overall accuracy of the Convolution Neural Network classification technique is performed. So this works gives better Classification results.

A. Calculation Formula

- TP: True positive (correctly predicted number of instances)
- FP: False positive (incorrectly predicted number of instances)
- 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

Total samples = 1150

Here it is found -

True Positive=933
False Positive=120
True Negative=790
False Negative=98

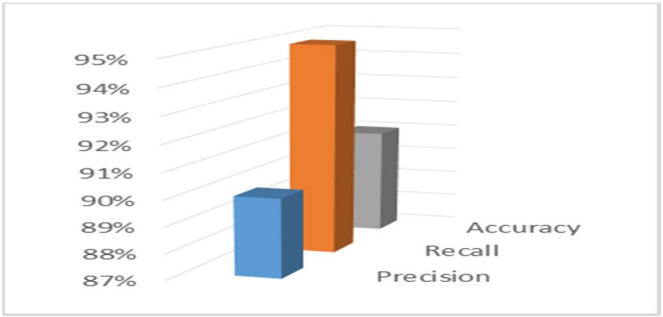


Fig-10- Performance Analysis Graph

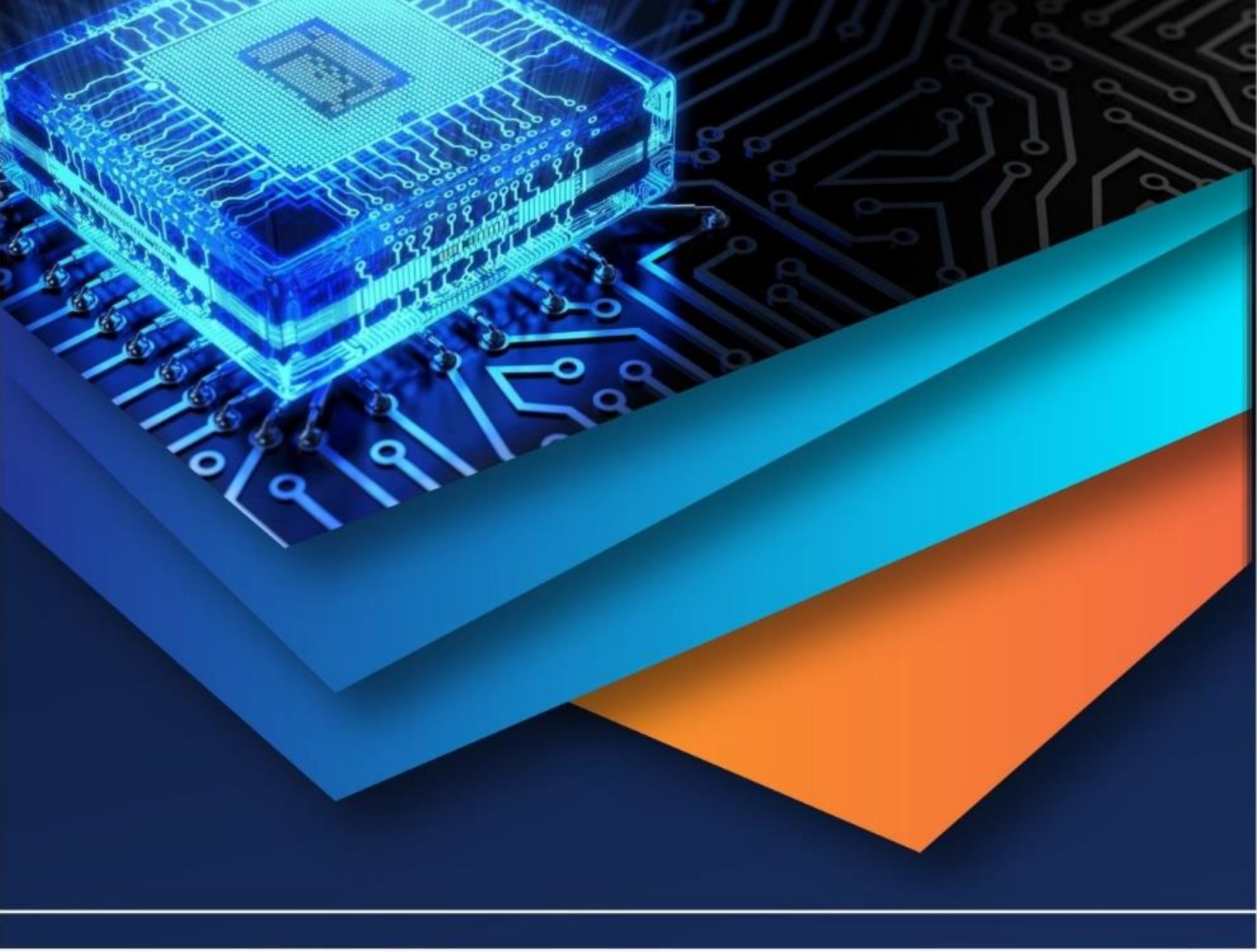
	Convolutional Neural Network
Precision	91.00%
Recall	90.20%
Accuracy	92.32%

VI. CONCLUSION

This proposed system presented an aspect-based sentiment classification framework that classifies reviews about aspects into positive or negative. In this framework, a tree-based aspects extraction method is proposed that extracts both explicit and implicit aspects from tourist opinions. It extracts frequent nouns and noun phrases from reviews text, and then groups similar nouns using WordNet. CNN is employed on reviews where review words are used as internal nodes and extracted nouns as the leaf of a tree. Opinion-less and irrelevant sentences are first removed by employing Stanford Basic Dependency on each sentence. Next, features are extracted from the remaining sentences with N-Grams and POS Tags to train the classifiers. Lastly, machine learning algorithms are applied to the extracted features to train the classifiers.

REFERENCES

[1] .M. Colhon, C. Badic˘a,˘ 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: 10.1007/978-3-319-12096-62
[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
[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



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)