10.48047/jocaaa.2024.33.08.378

Analysing Sentiments for Enriching Tourist Experience and Fostering Sustainable Growth in the Tourism Industry

Shanta H Biradar[a*], Dr. J V Gorabal [b], Dr. Swamy L N[c]

[a]Sir M Visvesvaraya Institute of Technology, Bengaluru.

[b]Department of Computer Science & Engineering, ATME College of Engineering, Mysore.

[c]Department of Computer Science & Engineering, Post Graduate Studies-Visvesvaraya Technological University,Mysore Region

[abc] Recognized by Visvesvaraya Technological University,Belgaum

***Corresponding Email:** shantha_is@sirmvit.edu[a*], jvgorabal@gmail.com[b], swamyln@gmail.com[c]

Abstract:

The study delves into the application of sentiment analysis models for evaluating and enhancing the tourist experience, with a specific emphasis on fostering sustainable growth within the context of Karnataka Tourism. While major social media platforms such as LinkedIn and Facebook have a significant impact on the industry, their focus primarily revolves around activities like hotel bookings and transportation reservations. However, there exists a distinct need for tools that can facilitate sentiment analysis of tourists' opinions, thereby allowing for the refinement of offerings based on valuable insights. The core objective of the research is to evaluate the quality of tourism experiences in Karnataka and raise them to international standards through the implementation of sentiment analysis techniques. The study utilizes data extracted from Twitter, spanning from August 1, 2022, to December 31, 2022. Emotional Analysis coupled with an Navies Bayes, Bernoulli naive Bayes, Multinomial naive Bayes, Support Vector Machine, Decision Tree and Romdom Forest deployed using the Python programming language are employed to analyse the data. The core objective of the research is to evaluate the quality of tourism experiences in Karnataka and raise them to international standards through the implementation of sentiment analysis techniques. The study utilizes data extracted from Twitter, spanning from August 1, 2022, to December 31, 2022. Emotional Analysis coupled with different classifier deployed using the Python programming language are employed to analyze the data. The labeled data, constituting 80% of the dataset, is utilized for the training of an different Classifier. The classifier's performance is then evaluated using the remaining 20% of the data, with the achieved accuracy reaching an impressive 84.596% (Navies Byes (MultinomialNaive Bayes). This methodology underscores the potential of sentiment analysis as a robust tool for assessing tourist sentiment, refining tourism offerings, and ultimately fostering both enriched tourist experiences and sustainable industry growth.

Keywords: Sentimental analysis, Artificial neural network, Tourist experience, twitter data

1. Introduction

The landscape of the tourism industry has undergone a radical transformation due to technological advancements related to the Internet, smartphones, and tablets [1]. This transformation has shifted the industry from a traditional brick-and-mortar and person-to-person service sector to a digitally empowered and omnipresent travel service network [2].This paradigm shift has granted individual travelers and groups an unprecedented level of control over their travel plans, enabling them to personalize and tailor their trips

10.48047/jocaaa.2024.33.08.378

according to their preferences.In parallel, the influence of online social media platforms such as Twitter, Instagram, Facebook, and others cannot be understated[3]. These platforms play a pivotal role in shaping electronic word-of-mouth (e-WOM), contributing to the amplification of travel-related opinions and experiences[4].However, with the rapid growth of online information, the task of extracting meaningful insights about tourists' attitudes, behaviors, and experience quality becomes increasingly complex and time-consuming. This challenge extends to both travelers and industry professionals, as well as researchers. In response, there's a growing demand for automated multi-aspect algorithms and machine-operated systems to handle the sheer volume of data effectively[5].

Many researchers[5-8] conducted research in particularly use of sentimental analysis in the development of various business such as hospital, hotel, educational institute etc. On the other hand researchers investigated the quality of tourist places using various theoretical model [6]. The tourism has many new challenges such as the transfer to newer type of international tourists, social media driven tourism and the use o new study technologies [7]. This study aims to analyze sentiments of Karnataka tourists by developing a hypothetical real-time comment system from social networks. It enhances model performance through improved sensitivity and specificity. It conducts a Comparative Analysis of data accuracy using Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest algorithms, focusing on Precision, Recall, and Accuracy. The research also explores user sentiment variations across different trends.

2. Related work

As previously introduced, the analysis centered on diverse tourism datasets. The results indicated that a significant proportion of sentiment analysis studies within the tourism field adopted a binary polarity classification approach. Several studies [8-9] employed a more nuanced strategy, involving a two-step process: initially categorizing sentences into objective and subjective classes, followed by determining polarity for subjective sentences. Moreover, certain investigations embraced a one-step technique, utilizing multiclass classifiers, particularly a 3-class classifier, to ascertain sentence polarity[10].

10.48047/jocaaa.2024.33.08.378

Nonetheless, comparability among outcomes is constrained due to variations in database sizes, class frequencies, and data types among evaluation methodologies. Notably, indications point towards superior results in terms of Accuracy and F-measure when restricting theanalysis to two classes (Positive, Negative)[11]. For equitable comparison of Twitter review outcomes, three methods were introduced alongside a publicly available Twitter-based product dataset for experimental purposes. Among these, Levallois's (2013) [12] method excelled in binary classification, while VADER (Hutto and Gilbert, 2014)[13] exhibited prominence in multi-class classification. Given their robust performance with Twitter data and the inclusion of tourism-related terminology and emoticons in their lexicons, these methods find practical application within the tourism domain[14].

The process of conceptualizing and modeling tourism facets involves a spectrum of variables. These variables encompass factors preceding the travel experience, those associated with recurring visits, electronic word-of-mouth (e-WoM), and loyalty. Scholarly literature [15-17] draws connections between objective quality and perceived quality, categorizing them into psychological, conative, and functional dimensions. Models that focus on tangible elements emphasize concrete aspects such as tourist attractions, facilities, and services. These models delve into visitor perceptions and prospective behaviors. On the other hand, alternative models classify tourist-destination experiences based on elements such as service, infrastructure, environmental aspects, and preservation of local resources. These models contribute to improved quality, value, and intentions for revisitation[18].

The fusion of perspectives from both tourists and residents holds pivotal importance within these models. Residents exert influence on tourism quality through a cost-benefit analysis encompassing event attachment and community concerns. Factors that shape this dynamic encompass overtourism, residents' quality of life, crowd levels, administrative quality, historical authenticity, and experience quality. Intangible variables, including psychological, behavioral, attitudinal, image, and brand quality, play a central role in shaping destination quality[19]. The sequence of image-quality-satisfaction serves as a significant foundation for understanding tourist satisfaction. This sequence underpins models related to destination loyalty, image perception, and behavioral intention. The framework of the experience economy underscores the outcomes of positive word-of-mouth (WoM), intentions for revisitation, and loyalty because of quality, satisfaction, and engagement[20].

With the integration of social networks and the proliferation of tourism engagement through platforms and e-commerce systems, the influence of brand quality, WoM, and e-

10.48047/jocaaa.2024.33.08.378

WoM expands[21].

10.48047/jocaaa.2024.33.08.378

This influence is evident through the impact of social media on cognitive aspects, affective imagery, brand quality, and customer valuation in shaping WoM and e-WoM. This trend promotes heightened tourist engagement and the intent to share travel intentions on interactive eReferral platforms, ultimately driving behavior related to eReferral sharing[22].

3. Developed Methods for classification of text

3.1 Classification of the text

Twitter text classification involves two main phases such as preparation and testing. The preparation phase encompasses creating named corpora datasets, preprocessing the training content, content vectorization, and classifier training. In the testing phase, testing content is pre-processed, vectorized, and classified.

3.2 Natural Language Processing

Short for natural language processing, NLP is focused on enabling computers to comprehend and interpret human language. Translating human language poses challenges due to its nuanced rules and contextual complexities.

3.3 Supervised ML Model to classify text

In the text classification process, a Twitter dataset with 80,000 rows of text is categorized into "Label 1" and "Label 2".

The dataset includes "text" and "Label" columns. To ensure reproducibility, a seed value is used, maintaining consistent results across script runs.

3.4 Setting Random seed.

Content is grouped based on classes, with each piece belonging to a specific class and labeled accordingly. The corpus is often divided into a Training set and a Testing set.

3.5 Creating corpus.

Unnecessary elements such as stop words, punctuation, and ambiguous content are removed from the text. This step is crucial to ensure effective classifier training.

3.6 Pre-Processing

Removed all unnecessary elements from the content, such as stop words, punctuation, or unclear text. This step is crucial because it will impact the training of the classifier.

3.7 Tokenization

Tokenization involves segmenting a stream of content into tokens, which is words, phrases, images, etc. Tokenization enables further processing, with NLTK Library offering tools like word tokenize and sent tokenize for word and sentence segmentation.

3.8 Word Stemming / Lemmatization

Both stemming and lemmatization aim to reduce inflected forms of words to their base or root forms. Lemmatization is closely related to stemming, but stemmers work on individual words without considering unique contexts, potentially lacking accuracy. However, they're

10.48047/jocaaa.2024.33.08.378

faster and easier to implement, with reduced precision potentially not impacting certain applications.

10.48047/jocaaa.2024.33.08.378

3.9 Vectorization of Text

Text Vectorization is a fundamental process that involves tokenizing a collection of text documents and building a vocabulary of recognized words. This process also facilitates encoding new documents using the established vocabulary.

3.10 Classifier Training

Select a text classification algorithm and provide the training corpus as input to the classifier, resulting in the creation of a training model.

4. Incorporating Sentiment Analysis Framework with an External System

The Open Authentication framework is a standardized method for obtaining credentials to access your information from various applications, ensuring that your login details are not shared with websites.

Third-party authentication grants access to protected information. Upon registering with Twitter to access the application interface, users are provided with a consumer account and a secret key. These credentials enable users to retrieve Twitter data.

Twitter data is publicly accessible, attainable either through web scraping or the Twitter API. The necessary credentials are obtained from the Twitter API to gather this data. Subsequently, an outcome variable must be established for these tweets, involving the classification of their sentiment as positive, negative, or neutral. The focus is on collecting tweets relevant to contemporary technologies like Blockchain, Machine Learning, Cybersecurity, and the Internet of Things, with the aim of incorporating them into the education system. The cleaning of data is a vital step. Stop words, often insignificant in many applications, are crucial to remove as they are commonly used words in documents. Words like "the," "a," "an," and "in" are disregarded. When indexing for queries, these words can clutter search results. The data is cleaned by eliminating numbers, spaces, special characters, etc. In this process, the 'gsub' function is employed.

4.1 Eliminate Stop Words, Non-Numeric Content, and Apply Word Stemming/Lemmatization

```
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk.tokenize import word_tokenize
# Download NLTK resources (run only once)
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')

def preprocess_text(text):
# Tokenize the text
words = word_tokenize(text)
# Remove punctuation and convert to lowercase
words = [word.lower() for word in words if word.isalpha()]
# Remove non-numeric words
words = [word for word in words if word.isnumeric()]
 # Stemming
 stemmer = PorterStemmer()
 words = [stemmer.stem(word) for word in words]
 # Lemmatization
lemmatizer = WordNetLemmatizer()
words = [lemmatizer.lemmatize(word) for word in words]
return words
```

10.48047/jocaaa.2024.33.08.378

The resultant set of words after processing for each iteration will be saved in the 'text_final' corpus. Storing will be done using the code: Corpus.loc[index, 'text_final'] = str(Final_words).

4.2 Create training and testing datasets.

The Corpus will be part into two informational collections, Training and Test. Thepreparation informational index will be utilized to fit the model and the forecasts willbe performed on the test informational collection. The Training Data will have 80%of the corpus and Test data will have the remaining 20% as we have set theparameter test size=0.3.The unprocessed data is categorized into different polarities: Label_1, Label_2, and Label_3, representing positive, neutral, and negative sentiments, respectively.

4.3 Word Vectorization

This involves a common approach of converting a collection of textual content files into numerical feature vectors. Several techniques exist for transforming textual data into comprehensible vectors by models, with the most widely used method being TF-IDF. This acronym stands for "Term Frequency Inverse Document Frequency," representing the values assigned to individual words. Term Frequency pertains to how often a specific word appears within a document, while Inverse Document Frequency diminishes the weight of frequently appearing words across multiple documents.

4.4 Executing the algorithm for the suggested model

4.4.1 Navies Bayes

The Naive Bayes Classifier technique relies on the Bayesian hypothesis and is particularly suitable when dealing with high-dimensional input. Despite its simplicity, Naive Bayes can often outperform more complex classification methods.

```
# Naive Bayes Classifier
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_vectorized,  y_train)

# Making predictions
predictions = nb_classifier.predict(X_test_vectorized)

# Evaluating accuracy
accuracy = accuracy_score(y_test, predictions)
```

4.4.2 Bernoulli naive Bayes

In the multivariate Bernoulli event model, features are represented as independent binary values (boolean variables) that define inputs.

10.48047/jocaaa.2024.33.08.378

Similar to the multinomial model, this approach is commonly employed in document representation tasks. Instead of using term frequencies (the occurrence count of a word in a document), the multivariate Bernoulli model uses binary term occurrence features.

4.4.3 Multinomial naive Bayes

Multinomial Naive Bayes is a specialized iteration of the Naive Bayes algorithm tailored for text documents. While the standard Naive Bayes approach considers a document's presence and absence of specific words, Multinomial Naive Bayes takes into account the actual word counts and adapts the underlying calculations accordingly.

```
# Multinomial Naive Bayes Classifier
multinomial_classifier = MultinomialNB()
multinomial_classifier.fit(X_train_vectorized, y_train)

# Making predictions
predictions = multinomial_classifier.predict(X_test_vectorized)

# Evaluating accuracy
accuracy = accuracy_score(y_test, predictions)
print("Accuracy:", accuracy)
```

4.4.3 Support Vector Machine

A Support Vector Machine (SVM) is a discriminative classifier characterized by a segregating hyperplane. When provided with labeled training data (supervised learning), the algorithm determines an optimal hyperplane that categorizes new instances. In a two-dimensional space, this hyperplane is a line that divides the plane into two sections, with each class positioned on one side. The objective of the support vector machine algorithm is to identify a hyperplane in an N-dimensional space (where N represents the number of features) that effectively separates the data points.

```
# Support Vector Machine Classifier
svm_classifier = SVC(kernel='linear')  # You can experiment with different kernels
svm_classifier.fit(X_train, y_train)

# Making predictions
predictions = svm_classifier.predict(X_test)

# Evaluating accuracy
accuracy = accuracy_score(y_test, predictions)
print("Accuracy:", accuracy)
```

10.48047/jocaaa.2024.33.08.378

4.4.4. Decision Tree

A Decision tree is a visual representation of possible outcomes resulting from a sequence of interconnected decisions. It enables an individual or an organization to compare potential actions by considering their costs, probabilities, and benefits.

```
# Training Decision Tree Classifier
dt_classifier = DecisionTreeClassifier(criterion='entropy')
dt_classifier.fit(X_train_vectorized, y_train)

# Making predictions
predictions = dt_classifier.predict(X_test_vectorized)

# Evaluating accuracy
accuracy = accuracy_score(y_test, predictions)
print("Accuracy:", accuracy)
```

4.4.5 Random Forest

Random Forest is a supervised classification algorithm that consists of a large number of individual decision trees functioning as an ensemble. This ensemble introduces two types of randomness. Firstly, each tree is built on a random sample from the original data. Secondly, at each tree node, a subset of features is randomly selected to determine the optimal split. Each individual tree within the random forest produces a class prediction, and the class with the highest count becomes the final prediction of our model.

```
# Training Random Forest Classifier
rf_classifier = RandomForestClassifier()
rf_classifier.fit(X_train_vectorized, y_train)

# Making predictions
predictions = rf_classifier.predict(X_test_vectorized)

# Evaluating accuracy
accuracy = accuracy_score(y_test, predictions)
print("Accuracy:", accuracy)
```

5. Experimental Results

The Table 1 provides the results of a sentiment analysis classification task using different machine learning models. Tourist sentiment analysis involves determining the sentiment or emotion expressed in a piece of text, typically categorized as positive, negative, or neutral.

Table 1. Comparative Performance Evaluation of Sentiment Analysis Classification Algorithms on a Specific Dataset

10.48047/jocaaa.2024.33.08.378

Model	Accuracy	Sensitivity	Specificity	Precision	F1-score
Navies Byes (**Binomial**	81.783	93.765	65.803	78.141	85.237
Naive Bayes)					
Navies Byes (**Multinomial Naive Bayes**)	84.596	93.765	72.377	81.536	87.221
Support Vector Machine(SVM)	76.137	74.025	78.960	82.247	78.812
Decision Tree (Gini Index)	65.014	59.220	78.960	78.960	67.679
Decision Tree (Entropy)	73.324	69.090	78.960	81.279	74.696
Random Forest	80.085	74.765	87.182	88.475	81.043

The models were evaluated based on several performance metrics such as accuracy, sensitivity, specificity, precision, and F1-score. Let's break down the meanings of these metrics and interpret the results for each model:Naive Bayes (Binomial Naive Bayes) model has high sensitivity and F1-score, indicating that it is effective at capturing positive instances. However, the specificity is relatively low, suggesting that it struggles with correctly identifying negative instances.**Naive Bayes (Multinomial Naive Bayes) which is same as** Naive Bayes model, this one has high sensitivity and F1-score. Additionally, it has a better specificity, making it a more balanced performer in terms of identifying both positive and negative cases. The SVM model has relatively balanced sensitivity and specificity, with an accuracy that falls in between the other models. It seems to perform reasonably well in both positive and negative class identification. This Decision Tree model, evaluated using the Gini Index criterion, has a lower accuracy compared to other models. Sensitivity and F1-score are also relatively low, indicating challenges in capturing both positive and negative instances effectively. The Decision Tree model using the Entropy criterion performs better than the Gini Index version. It has improved accuracy, sensitivity, and F1-score, making it more effective in capturing sentiment. The Random Forest model has high precision and specificity, indicating strong performance in correctly identifying both positive and negative instances. The F1-score is also decent, suggesting a good balance between precision and sensitivity. Finally different models exhibit varying strengths and weaknesses in sentiment analysis for tourist twitter data. The choice of the best model would depend on the specific goals and priorities of the analysis. Models like the Multinomial Naive Bayes and Random Forest seem to offer balanced performance across sensitivity, specificity, and accuracy for tourist twitter reviewer data. However, further optimization and tuning could be performed to enhance the performance of these models.

10.48047/jocaaa.2024.33.08.378

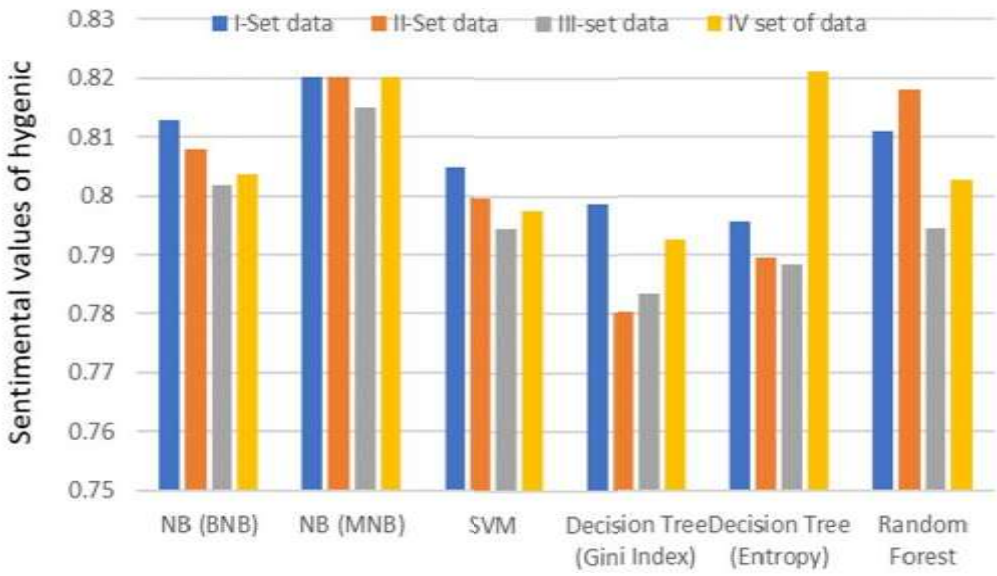

Fig. 1 Comparative Analysis of Sentiment Values for Hygienic Perception of Tourism Destinations across Different Datasets

The provided Fig. 1 showcases the results of a sentiment analysis conducted to assess the hygienic perception of tourism destinations using different classification algorithms. The sentiment analysis aims to quantify the sentiment values associated with the perception of cleanliness and hygiene across various sets of data. The algorithms employed include Naive Bayes (Binomial Naive Bayes and Multinomial Naive Bayes), Support Vector Machine (SVM), Decision Tree with Gini Index criterion, Decision Tree with Entropy criterion, and Random Forest.

The sentiment values, representing the hygienic perception scores, are computed for each algorithm and each dataset. Let's delve into the analysis of the results:

The sentiment values for Naive Bayes (BNB) span approximately from 0.8017 to 0.8129 across diverse datasets. This underscores the algorithm's consistent performance across the datasets, highlighting its robustness in effectively capturing perceptions related to hygiene. On the other hand, Naive Bayes (MNB) consistently maintains high sentiment values across all datasets, with a range between approximately 0.8149 and 0.8201. This unwavering pattern of sentiment values suggests that the algorithm excels in evaluating hygienic conditions with stable and positive outcomes. When observing Support Vector Machine (SVM), its sentiment values encompass a range of roughly 0.7946 to 0.8048 across the datasets. Although slightly lower than those achieved by Naive Bayes algorithms, SVM's sentiment analysis results exhibit a commendable consistency. In contrast, Decision Tree (Gini Index) portrays sentiment values ranging from about 0.7803 to 0.7925.

10.48047/jocaaa.2024.33.08.378

Notably, there's a discernible decrease in sentiment for the second dataset. This behavior points to varying sentiment values, indicating a moderate to lower perception of hygiene in that particular dataset. Similarly, Decision Tree (Entropy) showcases sentiment values fluctuating within the span of approximately 0.7885 to 0.8211. These shifts across different datasets reveal the algorithm's sensitivity to the characteristics of the data, leading to variations in sentiment analysis results. Moving to Random Forest, its sentiment values range approximately from 0.7946 to 0.8180. The consistency of these sentiment values underscores Random Forest's capacity to provide stable outcomes across datasets, signifying a balanced and effective performance in evaluating hygienic perceptions. Moving to Random Forest, its sentiment values range approximately from 0.7946 to 0.8180. The consistency of these sentiment values underscores Random Forest's capacity to provide stable outcomes across datasets, signifying a balanced and effective performance in evaluating hygienic perceptions.

6. Conclusion

In conclusion, the evaluation of diverse classification algorithms for sentiment analysis on tourist Twitter data has yielded insightful performance insights. Assessment metrics such as accuracy, sensitivity, specificity, precision, and F1-score revealed each algorithm's strengths and limitations.

- Naive Bayes (Binomial Naive Bayes) demonstrated high sensitivity and F1-score, excelling at capturing positive instances, yet struggling with negative identification accuracy.
- Naive Bayes (Multinomial Naive Bayes) shared these traits but exhibited better specificity, indicating a more balanced performance.
- The Support Vector Machine (SVM) displayed balanced sensitivity and specificity, with moderate accuracy, effectively identifying both sentiment classes.
- Decision Tree with Gini Index showed lower accuracy, sensitivity, and F1-score, indicating challenges in sentiment capture.
- Decision Tree with Entropy surpassed Gini Index, displaying enhanced sentiment capture abilities.
- Random Forest showcased high precision and specificity, competent in identifying both positive and negative instances.
- These findings underscore algorithmic nuances in sentiment analysis for tourist Twitter data. Multinomial Naive Bayes and Random Forest offer balanced performances.

10.48047/jocaaa.2024.33.08.378

- These insights inform algorithm selection for sentiment analysis, impacting the understanding of tourist perceptions and shaping strategies within the tourism industry.

References

1. Alireza Alaei, Susanne Becken, Bela Stantic (2017).Sentiment analysis in tourism: Capitalising on Big Data, Journal of Travel Research, 58(9):004728751774775.
2. Akehurst, G. (2009) "User generated content: the use of blogs for tourism organisations andtourism consumers." Service Business 3: 51-61.
3. Aurchana, P., R. PIyyappan, and P. Periyasamy (2014) "Sentiment analysis in tourism."International Journal of Innovative Science, Engineering & Technology 1 (9): 443-450
4. Manavi D Anicar, Dh Malini, (2018) Sentiment Analysis of eWOM on Service Performance of Public Undertaking Hotels, Amity Journal of Marketing, 3(1), (13–23)
5. Juan Pineda -Jaramillo, Martina Fazio, Michela Le Pira, Nadia Giuffrida, Giuseppe Inturri, Francesco Viti, Matteo Ignaccolo (2023). A sentiment analysis approach to investigate tourist satisfaction towards transport systems: the case of Mount Etna, Transportation Research Procedia, Vol. 69,(400-407).
6. Hall, C. M., & Ram, Y. (2019). Measuring the relationship between tourism and walkability? Walk Score and English tourist attractions. Journal of Sustainable Tourism, 27(2), 223-240
7. Inturri, G., Giuffrida, N., Le Pira, M., Fazio, M., & Ignaccolo, M. (2021). Linking public transport user satisfaction with service accessibility for sustainable mobility planning. ISPRS International Journal of Geo-Information, 10(4), 235.
8. Hall, C. M., & Ram, Y. (2019). Measuring the relationship between tourism and walkability? Walk Score and English tourist attractions. Journal of Sustainable Tourism, 27(2), 223-240
9. Choirul Huda, Yaya Heryadi, Lukas, Widodo Budiharto, (2019) Aspect-based Sentiment Analysis in Tourism Industry for Tourism Recommender System, *Sustainability*, *11*(18), 5070
10. Pablos, A.; Cuadros, M.; Linaza, M.T. OpeNER (2015): Open tools to perform natural language processing on accommodation. In *Information and Communication Technologies in Tourism*; Springer: New York, NY, USA, 2015
11. Xie, Z.H.; Wu, B.H. (2008) Tourism spatial structure of resources-based attractions in China. *Sci. Geogr. Sin.*, *28*, 748–753.
12. Levallois C. (2013) "Umigon: Sentiment Analysis for Tweets Based on Terms Lists and Heuristics." In Second Joint Conference on Lexical and Computational Semantics. Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation, 414-17
13. Hutto C., Gilbert E. (2014) "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, 216-25.
14. Kuttainen C., Lexhagen M., Fuchs M., Höpken W. (2012). "Social Media Monitoring and Analysis of a Swedish Tourism Destination." In Proceedings of the 2nd Conference on Advances in Hospitality and Tourism Marketing and Management
15. Sharma D., Kulshreshtha A., Paygude P. (2015) "Tourview: Sentiment Based Analysis on Tourist Domain." *International Journal of Computer Science and Information Technologies* 6 (3): 2318-20
16. Sigala M. (2016), "Social Media and the Co-creation of Tourism Experiences." In *The Handbook of Managing and Marketing Tourism Experiences*, edited by Sotiriadis M., Gursoy D., 85-111. Bingley: Emerald.

10.48047/jocaaa.2024.33.08.378

17. Veal A. J. (2006). *Research Methods for Leisure and Tourism: A Practical Guide*. 3rd ed. Harlow: Prentice Hall
18. *A García , S Gaines , M T Linaza ,A (2012) lexicon-based sentiment analysis retrieval system for tourism domain, Expert Syst Appl Int J , volume 39 , issue 10 , p. 9166 - 9180 Posted: 2012*
19. Z Abbasi-Moud , H Vahdat-Nejad , J Sadri (2021) Tourism recommendation system based on semantic clustering and sentiment analysis Expert Systems with Applications , volume 167 Posted,
20. S Aggarwal (2020) A Gour, Peeking inside the minds of tourists using a novel web analytics approach, Journal of Hospitality and Tourism Management , volume 45 , p. 580 - 591
21. S Ainin , A Feizollah , N B Anuar , N A Abdullah, (2020) Sentiment analyses of multilingual tweets on halal tourism, Tourism Management Perspectives , volume 34.
22. A De Maio , E Fersini , E Messina , F Santoro , A Violi (2020) Exploiting social data for tourism management: the SMARTCAL project, Quality & Quantity , p. 1 – 13.