International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

SENTIMENTAL ANALYSIS ON TOURISM REVIEWS

Ayalasomayajula Appala Narasimham[1], G Ajit Kumar[2], Goluguri V Manikanta Nagendra Reddy[3], G Kesava Anand Kumar[4], D Ravi[5]

1Assistant Professor, Computer Science and Engineering, Raghu Engineering College, Visakhapatnam

[2-5]B.Tech Students, Computer Science and Engineering, Raghu Institute Of Technology, Visakhapatnam

---***---

Abstract : *Sentiment analysis plays a pivotal role in understanding the sentiments and opinions expressed in textual data, offering valuable insights into various domains, including tourism. In this study, we present a comprehensive review of sentiment analysis techniques applied to tourism reviews using machine learning algorithms.*

The abundance of user-generated content on tourism platforms has made sentiment analysis an indispensable tool for businesses and researchers alike. By leveraging machine learning algorithms, researchers can extract sentiments from vast amounts of textual data efficiently and accurately.

This review outlines the key methodologies and approaches utilized in sentiment analysis of tourism reviews. It discusses preprocessing techniques such as text tokenization, stop-word removal, and stemming, which are crucial for preparing textual data for analysis. Furthermore, it examines various machine learning algorithms employed for sentiment classification, including Naive Bayes, Support Vector Machines, and Recurrent Neural Networks.

Additionally, the review delves into feature extraction methods such as bag-of-words, TF-IDF, and word embeddings, highlighting their impact on sentiment analysis accuracy. Moreover, it explores the challenges and limitations associated with sentiment analysis in the tourism domain, such as sarcasm detection and language nuances.

I.INTRODUCTION :

In recent years, the tourism industry has witnessed a significant transformation driven by the proliferation of online platforms and social media channels. These platforms have become a breeding ground for user-generated content, including reviews, feedback, and recommendations. As a result, analyzing the sentiments expressed in tourism reviews has become crucial for businesses and researchers alike to understand customer preferences, improve service quality, and enhance overall customer experiences

Sentiment analysis, also known as opinion mining, is a computational technique that involves extracting subjective information from textual data to determine the sentiment or opinion conveyed by the author.

In the context of tourism, sentiment analysis provides valuable insights into the experiences, perceptions, and emotions of travelers, helping tourism businesses and destinations make data-driven decisions

The primary objective of this project is to develop a sentiment analysis system tailored specifically for tourism reviews. By employing machine learning algorithms and natural language processing techniques, we aim to analyze the sentiments expressed in tourism reviews to identify positive, negative, and neutral opinions.

This project will contribute to the existing body of knowledge in several ways:

Improving Customer Satisfaction: By understanding the sentiments expressed in tourism reviews, businesses can identify areas for improvement and address customer concerns more effectively, ultimately enhancing overall customer satisfaction.

Enhancing Reputation Management: Tourism destinations and businesses can use sentiment analysis to monitor their online reputation and respond promptly to negative reviews, thereby mitigating potential damage to their brand image.

Informing Marketing Strategies: Sentiment analysis can provide insights into the preferences and interests of travelers, enabling tourism businesses to tailor their marketing strategies and promotional campaigns accordingly.

Supporting Decision-Making: Data-driven insights derived from sentiment analysis can inform strategic decision-making processes within the tourism industry, such as product development, pricing strategies, and resource allocation.

In the following sections of this project, we will delve into the methodologies, algorithms, and techniques employed in sentiment analysis of tourism reviews. We will explore the process of data collection, preprocessing, feature extraction, and sentiment classification. Additionally, we will discuss the implementation of a prototype sentiment analysis system and evaluate its performance using real-world tourism review datasets.

Overall, this project aims to demonstrate the utility and effectiveness of sentiment analysis in extracting valuable insights from tourism reviews, thereby facilitating informed decision-making and driving continuous improvement within the tourism industry.

II. LITERATURE SURVEY:

According to the changes of technology which is related to the Internet, including smartphones and tablets, have remodeled the tourism industry from a stuff and direct service industry to a stiffy digitally supported and universal travel service network. To increase their idea in reference to traveling and decision making in tourism, they not only collaborate with a scope of program and online agent however they also associate with other tourists who share their experiences. Tourists who have access to online platforms to give their feedback and make approvals for other tourists [3].

To identify about the strengths and defects of different products and services we need customer generated contents that helped consumers, and find the ones that give the best fit for their requirements. However, customer generated contents due to their volume, variety, velocity and veracity, establish huge challenge for businesses as well as for customers in analyzing and deriving insights from them [4]. Sentiment analysis tries to spot and analyze opinion and emotions [5]. Sentiment could be an equivalent word of attitude in tourism research. The attitudes of each tourists and residents play vital roles in tourism development, since friendly interactions between guests and residents produce great impact on visitors' satisfaction [6]. Recent tourism literature has studied problems with place identity in understanding resident attitudes toward tourism [7, 8, 9, 10,11]. These studies conclude that local people's sentiment is crucial to the sustainability of tourism development.

The sentiments of tourists are vital for the development of touristy.In [12], on the web acceptance of restaurants, the researchers experiment to recognize the impact of unsubstantiated information. Tourists' online reviews can also guide people to decide on travel agency and destination. Since the applying of sentiment analysis in tourism has become more and more popular. With machine learning approaches, most researchers study the sentiment classification in tourism. In [13], in order to automatically classify the customer reviews as positive, negative or neutral, standard machine learning techniques naive Bayes and SVM are integrated. In [14], Zheng and Ye conduct an exploring analysis on sentiment analysis to Chinese traveler reviews by SVM algorithm. Compared to review of machine learning approaches, in tourism, there are only a few papers study the lexicon-based

March - April 2020 ISSN: 0193-4120 Page No. 2466 - 2474 2468 Published by: The Mattingley Publishing Co., Inc. approach. for example, Kang, Yoo and han [15] propose a replacement seniti-lexicon for the sentiment analysis of building reviews. In [16], tourism will increase employment opportunities for the local individuals, improves the local economy, contributes to income and customary of living, brings in new businesses and improves investment opportunities. From the advertisement of customer reviews, opinion mining or opinion extraction, has evolved into rebellious in digital marketing world for attribution aspect of Internet.

Those reviews which are being actual time interpretation of the customer emotions, those customers written text are relatively more reliable and genuine than any other source of text data.[17] On industry of tourism and hospitality, transportation of air, framework development and progress in technical has direct influence. Automated sentiment analysis which has been used for extracting opinions from public and various other sourcesfor different applications in tourism, marketing and hospitality comparing with their performance to that of the human evaluatorsto estimatesthe compatibility of various types of automated classifier.[18] There is a proposed framework of opinion mining in tourism for summarizing visitors' opinions and experiences from TripAdvisor reviews. This framework is used to classified the reviews into multiple categories and performs lexicon-based sentiment extraction in each category. [19][20] In [21], Feldman propose five main problems for

III.METHODOLOGY

Sentiment analysis and data/information visualization methodologies guide this research. Sentiment analysis methodology: In the first part of my research, sentiment analysis methodology is the means to convert unstructured data to information. There are specific steps to analyze sentiment data as seen in Figure 1, namely: 1. Data collection 2. Text preparation 3. Sentiment detection 4. Sentiment classification 5. Presentation of output Figure 1: Sentiment analysis methodology The first step is data collection. This process is either based on a database or it is a real time collection from public forums, blogs and social network sites. The gathered elements are stored and then processed (text preparation) because opinions and feelings are expressed in ways that are making the data huge and disorganized. Text preparation is filtering the data before analysis through elimination of non-textual or irrelevant content. Typical pre-processing procedure includes tokenization, stemming, lemmatization, stop-words removal and lowering the case of the letters [30]. The third step pertains to sentiment detection and examines the subjectivity of each element which leads to the sentiment classification. During sentiment classification, each subjective sentence detected is grouped by its polarity (negative, neutral, positive). The final stage is to present the output. The meaningful information extracted by this process is displayed on graphs and diagrams. The recommended programming languages to perform Sentiment Analysis is Python and R, an open source tool for statistical computing and graphics. In this case, I have chosen to work in the Python environment. Python is suitable for text processing while R is preferred when data is in other forms. Moreover, Python has a variety of libraries and small supporting tools. Most of the researchers working in sentiment analysis use Python, therefore there is an active community as well. In conclusion, Python is a familiar tool and my personal preference. Data/Information visualization: The approach of designing visualizations can be divided into six steps:

1. mapping
2. selection
3. presentation
4. interactivity
5. usability
6. evaluation

The first step is mapping and it determines how to encode information into visual graphics. Mapping is the most efficient translation of data objects into visual objects. This process contributes to the understanding of the offered information and prepares the ground for the next step. Selection means choosing the appropriate data among every element according to the given task. It is essential to define what questions should be answered and then, select the proper data for this. After mapping and accurate selection of data it is important to present the information in the most suitable and understandable form (presentation stage). According to Tufte, the most important characteristics of an effective graphical form is accuracy, several levels of details, cohesion and clarity [31]. The right decision in terms of presentation allows greater interactivity. User friendly interactivity gives the user the opportunity to explore the information easily. The next step in the visualization process is usability, namely taking into consideration human factors. The graphic plots are addressed to people from different backgrounds, so the visualization must be universally comprehended. The final step is the evaluation of the created form to find out whether it was effective or not. In this project, evaluation forms an aspect of future work.

IV. IMPLEMENTATION

This chapter describes the implementation of sentiment analysis on the collected data set, as well as the visualization approach taken. In particular, the process of implementation presented below includes data collection, data preprocessing, the application of algorithms onto the transformed data, and data visualization.

DATA COLLECTION: The first step in the process of sentiment analysis is the "data collection". In my case, the collected data pertain to traveling experiences and tourist reviews pertaining to Athens. The main concerns are the origins of the data and the types of entities.

ORIGINS OF DATA:

- Twitter
- Airbnb
- FourSquare
- Google Maps

Types Of Entities:

The selection of the entities to collect comments/reviews took place according to the criteria that define whether a city has been touristically developed. The received information is divided into categories (entertainment, museum, food, nightlife), despite Twitter posts which have no category and Airbnb data that are classified as accommodation.

Field	Explanation
id	Original post's identifier. It operates as a reference to the json file that contains the particular post; each filename consists of the retrieved post's id.
origin	The origin of the post. It is essential for statistical reasons
date	The date that the post was written, if it is included (optional).
review_text	The main information of the posts, which is processed and classified according to its sentiment.
clean_text	The review's text after the data preprocessing.
rating	The original text's rating, if it is included (optional).
location_lat	The location's latitude of the related
location_long	The location's longitude of the related review.
venue_category	The specific category of the venue (i.e. apartment, theater, bar) (optional).
type	The general type of the described venue. (i.e. entertainment, food) (optional).
topic	topic Three keywords that specify the text's topic
address	The location of the related review.
sentiment	The sentiment (negative = -1.0 , positive = 1.0, neutral = 0.0) as occurred from the sentiment analysis algorithm or the given rating

Data preprocessing: A significant part of the collected data are gibberish or not relevant to the topic of the project but cannot be eliminated during data collection. Data preprocessing takes place in order to eliminate the irrelevant elements that could possibly alter the result of the research. By the end of this process, the data become compatible input for the data analysis algorithms.

1.Language detection :The Airbnb dataset contained an amount of non-english reviews which needed to be excluded from the preprocessed data. Using detect function from the langdetect library, 14 each review that was not exclusively english text was deleted from the set.

2. Cleaning reviews/posts: The reviews' and posts' body includes many words and characters that are impossible to analyze at an algorithmic level and, consequently, should be removed. Links, stopwords, punctuation symbols, emojis are ignored and the remaining text is transformed into lowercase. For example, after the cleaning procedure, the review's text "Amazing places to visit !!! http://… " is converted into "amazing places to visit".

3. Tokenization: Tokenization is the process of demarcating and possibly classifying sections of a string of input characters. Tokenization constitutes an essential step in lexical analysis. Texts need to be converted from uniform strings into tokens because each word will be transformed into a vector in order to become an acceptable input for machine learning algorithms (Image 5). The division into multiple tokens is, moreover, appropriate for lexical statistics (topic modeling, calculation of words' frequency, creation of word clouds etc.). In Python, tokenization is implemented via modules of nltk library. Using the above 15 examples, "amazing places to visit" is transformed into an array with content ["amazing", "places", "to", "visit"].

4.Lemmatization: In computational linguistics, lemmatization is the algorithmic process of determining the lemma of a word based on its intended meaning. Lemmatization for the English language is relatively simple but not a manual process, because hand-written lemmatization rules can be troublesome. Nltk library uses machine learning techniques to help developers with the stemming process. The algorithms' input does not include only words. Additionally, it includes metadata such as POS tag. POS tagging labels the words according to a set of tags, known as tagset, that usually contains parts of speech (noun, adverb, verb etc.). This procedure helps to determine the correct lemma for ambiguous forms.

V. ALGORITHM APPLICATION

When the data preprocessing procedure is completed, data are ready to be transformed into an appropriate format for algorithms application. Topic modeling and sentiment analysis algorithms contribute to the automatic extraction of information, taking into consideration both the diction and the content of each post/review

Custom classifiers :The sentiment analysis task is usually modeled as a classification problem where a classifier is fed with a text and returns the corresponding category, e.g. positive, negative, or neutral (in case polarity analysis is being performed). The first step is the training process. During this stage, the model associates a particular input to a specific output based on the test samples. The training is followed by the prediction process. In the prediction process, the feature extractor transforms unpredicted text inputs into feature vectors. These vectors are reprocessed and generate the predicted tags (positive, negative, neutral).

The model's input requires the text to be transformed into a numerical representation (vector). This process is known as 'feature extraction' or 'text vectorization'. There are many different algorithms that are suitable for this procedure e.g. BOW, TF-IDF, bag-of-ngrams. New feature extraction techniques include word embeddings, known as word vectors. This kind of representation makes it possible for words with similar meaning to have a similar representation, which could generally improve the performance of classifiers.

- **BOW:** A bag-of-words is a representation of a text that describes the occurrence of words within a document. The main involved notions are a vocabulary of known words and a measure of the presence of known words. As implied by its name, bag-of-words does not take into account any information about the order or structure of words. The only concern is the word itself and not its location in the document. As a result, documents are similar if they have similar content.
- **TF-IDF** TF-IDF is a metric that represents how 'important' a word is to a document. Scoring word frequency has an occupational hazard. Highly frequent words start to dominate in the document (e.g. larger score), but may not contain as much "informational value" to the model as rarer words. The vectorizer's concept is to monitor the number of times a word appears in a document taking into account the overall appearance in every document. Words that are generally common rank low, even though it may occur many times, since it is implied that they do not have importance to that particular document [5]. TF-IDF means Term Frequency - Inverse Document Frequency. Thus, the idf of a rare term is high, whereas the idf of a frequent term is low.

The classification stage usually involves a statistical model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks. Unlike traditional probabilistic and non-probabilistic algorithms, neural networks are associated with deep learning and consist of a diverse set of algorithms that attempts to imitate the brain's functionality by employing artificial networks to process information.

- **SVM**: Support vector machines are widely used in classification tasks. Their main data structure is a hyperplane. Hyperplanes are decision boundaries that help classify the data points. Depending on the side of the hyperplane that the data points fall, it is decided in which class they will be assigned [8]. The input of SVM is called sample features. It is a training set of data with predefined sentiment. Training data are vectorized and afterwards, are inserted into SVM paired with the corresponding sentiment label.

In this project, the training set for classifying the posts of each online platform differed. Moreover, the lookup for well trained open source data was a difficult task. Trained data for Twitter are 28061 22 tweets that are obtained from a university research within the framework of SemEval content in 2017. Regarding foursquare, the dataset is composed of tips 23 referring to the localities of the city of São Paulo/Brazil. The tips belong to the foursquare's categories: Food, Shop & Service and Nightlife Spot and contain 179181 tips. Before being inserted into the vectorizer, the tips were translated into english to achieve more accurate vectors. In general, a training set is preferred to have the same content as the test set and to contain a larger amount of information.

Twitter and Foursquare data were processed via an SVM classifier, because datasets for Airbnb and Google Maps were unavailable. Google Maps reviews' sentiment has occurred via rating. Airbnb sentiment analysis was only handled by an automatic data analysis tool which is discussed below. The output of the SVM includes a set of weights, one for each feature, whose linear combination predicts the value of the test set. The test set in this project is formed by the vectorized collected posts and reviews from online platforms.

After the end of the classification stage, each post is matched with a predicted sentiment (-1 if negative, 0 if neutral, 1 if positive). The modules needed for SVM classification were imported from sklearn package. During the evaluation stage the classifier's performance metrics are obtained, in an attempt to comprehend the accuracy of the sentiment analysis model. The most frequent evaluation method is cross-validation.
The training dataset is divided into a certain number of training folds and testing folds. The training folds are used to train the classifier and the testing folds are used to test the classifier and obtain performance metrics. The process is repeated multiple times, until an average of each metric is calculated. The standard metrics to evaluate a classifier are precision, recall and accuracy. Precision measures the percentage of the correctly predicted texts of a particular category. Recall measures the percentage of the correctly predicted texts of a particular category, given the texts that should have been predicted in the specific category.

This metric will be improved if the classifiers' input data are increased. Accuracy measures the percentage of the correctly predicted texts out of all the texts in the corpus. For a difficult task like analyzing sentiment, precision and recall levels are likely to be low at first and will increase as more data are given. If the testing set is always the same, it might lead to overfitting. Overfitting means over adjusting the analysis to a given dataset and consequently failing into analyzing any other dataset. Cross-validation contributes to preventing that phenomenon. In this project, the classifier' s evaluation was not a main desideratum, however the results proved that overfitting was avoided. In case of overfitting the predicted values would be tampered, hence useless.

- **K-Nearest Neighbor Classifier:** In order to use a classifier within this approach, KNN is selected. Since, sentiment analysis is a binary classification and there are huge datasets which can be executed, KNN is chosen here. A manually generated training set is utilized for training the classifier here. There is X:Y relation provided within the training set in which the score of an opinion word is represented by x and the score whether the word is positive or negative is represented by y [15]. A score of the opinion word related to a feature within the review is given as input to KNN classifier.

All the reviews that include that feature are to be considered in order to extract the opinion relevant to a particular feature. The ratio of total number of reviews that include a positive sentiment to the total number of reviews given is computed as the eventual positive score for particular feature. The ratio of total number of reviews within which a negative sentiment related to a feature is given to the total number of reviews present is calculated as the eventual negative score for particular feature.

VI. RESULTS AND ANALYSIS

The sentiment analysis and data/information visualization methodologies followed are popular and widely used processes. However, each stage lurked its own challenges. During the collection of reviews , finding a decent amount of valid open source data was a slow and difficult process that required time and manual parameter entry on a daily basis; hence, it could not be fully automated. Next, raw data have undergone a series of transformations in order to be compatible input for the machine learning algorithms and the visualization libraries. The data preprocessing stage, in this project, could be labeled as an agile procedure. Questions' increase causes further data preprocessing in order to extract the necessary answers. Machine learning algorithms included both custom classifiers and ready-made tools. The implementation fulfilled the hypothesis of combining different methods to achieve higher accuracy.

However there is space for improvement in case more data or a better trained dataset is collected. Moreover, the posts that have been labeled as neutral, which was the great majority, could have been dealt with in more detail, taking into consideration sarcasm or opinion words. It was fascinating to see how the lexicons and the training datasets are probably the most important piece in a sentiment analysis system.

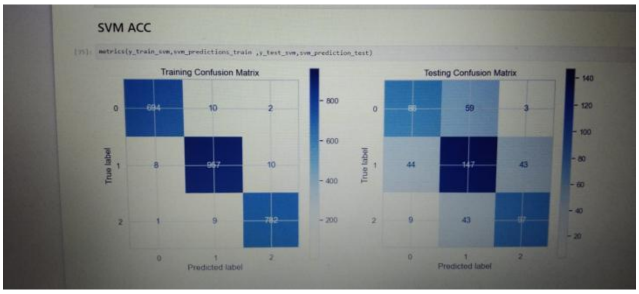
To analyze the performance of the proposed system various performance analysis metrics are considered like precision, recall and accuracy. The performance of the proposed system is compared with the existing system in which SVM classifier is used for the classification of positive, negative and neural tweets. The formula of precision is given by equation 1, recall is defined with equation 2 and accuracy is defined by equation by formula 3

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \text{ ------------------ (1)}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \text{ --------------------- (2)}$$

$$\text{Accuracy} = \frac{\text{No of tweets correctly classified}}{\text{Total Number of tweets}} \text{ ------------- (3)}$$

Model	Training Accuracy	Testing Accuracy	Validation Accuracy
SVM	91%	85%	85%
KNN	75%	74%	74%



VII.CONCLUSION

In this paper, the sentiment analysis system is presented which is based on N-gram and KNN classifier. In the past years various techniques are designed for the sentiment analysis. The proposed system is inspired from the technique in which SVM classifier is used for the classification of positive, negative and neural tweets. The proposed system is based on N-gram and KNN classifier. The features of the input data are extracted with N-gram algorithm and KNN classifier is applied to classify data into positive, negative and neural classes. The performance of proposed system is

compared with existing SVM classifier system. The experimental result shows upto 7 percent improvement of sentiment analysis. The experiments are conducted on the English data and in future performance of proposed system can be tested on other languages.

VIII. REFERENCES

[1] "Choropleth map," wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Choropleth_map. [Accessed: 04-Jan-2020]. [2] A. VLACHOU, "Greece's coffee industry grows despite financial crisis," ekathimerini, 2018. [Online]. Available: http://www.ekathimerini.com/230021/article/ekathimerini/business/greeces-coffee-industry-gro ws-despite-financial-crisis. [Accessed: 13-Dec-2019]. [3] "Athens is the world's ancient capital," lonely planet. [Online]. Available: https://www.lonelyplanet.com/greece/athens. [Accessed: 12-Dec-2019]. [4] "Athens," urbact. [Online]. Available: https://urbact.eu/athens. [Accessed: 13-Dec-2019]. [5] B. Stecanella, "What is TF-IDF?," MonkeyLearn, 2019. [Online]. Available: https://monkeylearn.com/blog/what-is-tf-idf/. [Accessed: 07-Dec-2019]. [6] Jason Brownlee, "A Gentle Introduction to the Bag-of-Words Model," Machine Learning Mastery, 2017. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-bag-words-model/. [Accessed: 07-Dec-2019]. [7] S. Fan, "Understanding Word2Vec and Doc2Vec," shuzhanfan.github.io, 2018. [Online]. Available: https://shuzhanfan.github.io/2018/08/understanding-word2vec-and-doc2vec/. [Accessed: 07-Dec-2019]. [8] R. Gandhi, "Support Vector Machine — Introduction to Machine Learning Algorithms," towardsdatascience, 2018. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algo rithms-934a444fca47. [Accessed: 09-Dec-2019]. [9] J. Brownlee, "What Are Word Embeddings for Text?," Machine Learning Mastery, 2017. [Online]. Available: https://machinelearningmastery.com/what-are-word-embeddings/. [Accessed: 07-Dec-2019]. [10] E. L. Steven Bird, Ewan Klein, "Categorizing and Tagging Words," in Natural Language Processing with Python, O'Reilly Media, 2009, p. 504. [11] "Topic model," wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Topic_model. [Accessed: 18-Nov-2019]. [12] "Όρια Συνοικιών Δήμου Αθηναίων," geodata.gov.gr, 2019. [Online]. Available: http://geodata.gov.gr/el/dataset/op1a-euvo1k1wv. [Accessed: 18-Nov-2019]. [13] P. Pandey, "Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)," medium, 2018. [Online]. Available: https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-i n-python-f9e6ec6fc52f. [Accessed: 22-Nov-2019]. [14] "Topic Modeling with Gensim (Python)," machinelearningplus. [Online]. Available: https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/. [Accessed: 18-Nov-2019]. [15] "Lemmatisation." [Online]. Available: https://en.wikipedia.org/wiki/Lemmatisation. [16] E. Fonseca, "State-of-the-art Multilingual Lemmatization," towardsdatascience. [Online]. Available: https://towardsdatascience.com/state-of-the-art-multilingual-lemmatization-f303e8ff1a8. [Accessed: 13-Nov-2019]. [17] "Lexical analysis." [Online]. Available: https://en.wikipedia.org/wiki/Lexical_analysis#Tokenization. [Accessed: 12-Nov-2019]. [18] P. Gil, "What Is Twitter & How Does It Work?," lifewire, 2019. [Online]. Available: https://www.lifewire.com/what-exactly-is-twitter-2483331. [Accessed: 28-Oct-2019]. [19] "Place Search," Google Maps Platform. [Online]. Available: https://developers.google.com/places/web-service/search. [Accessed: 29-Oct-2019]. [20] "Place Details," Google Maps Platform. [Online]. Available: https://developers.google.com/places/web-service/details. [Accessed: 29-Oct-2019]. [21] A. Ide, "A Turning Point for Tourism Informatics," New Breeze, vol. 29, no. 4, p. 4, 2017. [22]. M. Roussou et al., "Deliverable D4.1 - Indicators visualization and representation modelling and techniques," enventory - European eInfrastructures Observatory project report, 2011. [23] E. Bruin, "Airbnb: The Amsterdam story with interactive maps," kaggle, 2019. [Online]. Available: https://www.kaggle.com/erikbruin/airbnb-the-amsterdam-story-with-interactive-maps. [Accessed: 21-Oct-2019]. [24] "Top 7 Data Science Use Cases in Travel," ActiveWizards. [Online]. Available: https://activewizards.com/blog/top-7-data-

science-use-cases-in-travel/. [Accessed: 21-Oct-2019]. [25] M. Aggarwal, "Application Of Machine Learning and Deep Learning In the Hospitality Industry," medium, 2018. [Online]. Available: https://medium.com/@manuj.aggarwal/application-of-machine-learning-and-deep-learning-in-the-hospitality-industry-ca9675ce7b94. [Accessed: 21-Oct-2019]. [26] A. Bulanov, "Benefits of the Use of Machine Learning and AI in the Travel Industry," djangostars. [Online]. Available: https://djangostars.com/blog/benefits-of-the-use-of-machine-learning-and-ai-in-the-travel-industry/. [27] M. Khan and S. S. Khan, "Data and information visualization methods, and interactive mechanisms: A survey," Int. J. Comput. Appl., vol. 34, no. 1, pp. 1–14, 2011. [28] GauravChhabra, "NLP - Twitter Sentiment Analysis Project," kaggle, 2019. [Online]. Available: https://www.kaggle.com/gauravchhabra/nlp-twitter-sentiment-analysis-project. [Accessed: 21-Oct-2019]. [29] H. M, "TripAdvisor Reviews - Data Analysis," kaggle, 2019. [Online]. Available: https://www.kaggle.com/harimo/tripadvisor-reviews-data-analysis. [Accessed: 21-Oct-2019]. [30] O. Kolchyna, T. T. P. Souza, P. C. Treleaven, and T. Aste, "Methodology for Twitter Sentiment Analysis," arXiv Prepr. arXiv1507.00955, 2015. [31] E. R. Tufte, The visual display of quantitative information, vol. 2. Graphics press Cheshire, CT, 2001. [32] A. Mattaranz, "Applications of Text Analytics in the Tourism Industry," meaning cloud, 2017. [Online]. Available: https://www.meaningcloud.com/blog/applications-for-text-analytics-in-the-tourism. [Accessed: 21-Oct-2019]. [33] Gaurav Shankhdhar, "Sentiment Analysis Methodology," edureka, 2019. [Online]. Available: https://www.edureka.co/blog/sentiment-analysis-methodology/. [Accessed: 20-Oct-2019]. [34] "Sentiment Analysis The Only Guide You'll Ever Need," MonkeyLearn. [Online]. Available: https://monkeylearn.com/sentiment-analysis/. [35] "Project," Public participation city. [Online]. Available: https://ppcity.eu/. [Accessed: 20-Oct-2019]. [36] "Tourism Sentiment Index," destinationthink. [Online]. Available: https://destinationthink.com/about-tsi/. [Accessed: 21-Oct-2019]. [37] J. Wu, "AI, Machine Learning, Deep Learning Explained Simply," medium, 2019. [Online]. Available: https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960. [Accessed: 21-Oct-2019]. [38] B. Bettendorf, "NLP on Airbnb Data," kaggle, 2019. [Online]. Available: https://www.kaggle.com/brittabettendorf/nlp-on-airbnb-data. [Accessed: 21-Oct-2019].