Sentiment Analysis towards Hotel Reviews

*Nur Hidayu Md Nohh, Norziha Megat Mohd. Zainuddin,
Syahid Anuar, Nurulhuda Firdaus Mohd Azmi,
Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia, 54100 Kuala Lumpur
nurhidayumdnoh92@gmail.com, norziha.kl@utm.my,
syahid.anwar@utm.my, huda@utm.my,

Wan Azlan Wan Hassan
Faculty of Communication Visual Art and Computing
Universiti Selangor, Bestari Jaya, 45600 Selangor
wan.azlan@unisel.edu.my

Article history

Received:
9 Sept 2019

Received in revised form:
31 Oct 2019

Accepted:
20 Dec 2019

Published online:
30 Dec 2019

*Corresponding author
nurhidayumdnoh92@gmail.com

Abstract

The used of social media analytic towards the information in social media could bring a significant impact toward the hotel business thus making it easier to gain such insight and analysis. A preliminary study showed that the hoteliers in Malaysia are unaware on the sentiment analysis computation in analyzing the customer review through TripAdvisor despite having the website to promote their hotel. This study aims to provide another addition of sentiment analysis study especially on hotel reviews dataset gain from TripAdvisor. The process start by scrapping the reviews from The Majestic Hotel in TripAdvisor. The dataset then undergoes a pre-processing stage using Valence Aware Dictionary for Sentiment Reasoning to identify the sentiment in the text reviews. The text reviews are converted into vector form and fed into machine learning classifier. Three type of machine learning classifier was chosen to classify the review that are Naïve Bayes, Support Vector Machine and Maximum Entropy. The classifiers are then evaluated using three measurement from the accuracy, precision and recall value perspectives in order to find the best classifier. Based on this study, Maximum Entropy gained the highest value compared to the other two classifier.

Keywords: *Sentiment Analysis, Hotel Reviews, Social Media Analytics, Machine Learning*

1. Introduction

Hotel organizations worldwide are gaining huge and valuable information and insight towards their product and services by applying appropriate tools towards their data such as their hotel reviews. This reviews will generates a huge volumes of information known as big data (Jimenez-marquez *et al.*, 2019). Nowadays people tends to search reviews of product or services before they decided to use the products or services (Kaplan and Haenlein, 2010). This is due to people dependencies towards other's experience first before making a decision.

In hospitality industry, the customers usually shared their experience through comments and reviews on the hotel services. The progression of social media encourage the customers to give those related recommendation on a collaborative platform. The feedback from the customers is significant to enhance the services offered in the future. The data from the reviews contain a lot more information that can be manipulated in order to increase hotel performances. Conducting a manual analysis or survey using the dataset is costly and only limited data could be obtained and utilized (Chang *et al.*, 2017; Mariani *et al.*, 2019). Thus, by utilizing social media analytics, the collection, monitoring and analyzing the reviews could be done using advance information tools and analytics techniques (He *et al.*, 2015). Therefore, the technique and process involved in social media analytics comprise of the data analytics, text mining and sentiment analysis.

The use of social media analytics can bring a significant impact towards hospitality industry. However, the research mostly focus on sentiment analysis in reviewing Twitter and Amazon while the sentiment analysis on hotel reviews dataset is still lacking (Geetha *et al.*, 2017). Thus, this study aim to leverage the sentiment analysis on reviewing the hotels dataset. This study utilized several machine learning algorithm such as Naïve Bayes, Support Vector Machine (SVM) and Maximum Entropy algorithm.

This study is to achieve the following objectives:

- To identify sentiment analysis study on hotel reviews dataset.
- To develop a corpus sentiment analysis towards hotel reviews by using machine learning classifier.
- To evaluate which machine learning classification in corpus sentiment analysis perform the best in classifying the reviews.

This article is organized into several section. Section 1 shall give the introductory and the aims of the article. Section 2 contain the literature reviews related to the sentiment analysis fields which include social media analytics, sentiment analysis and machine learning. Section 3 briefly presents related work on sentiment analysis using different kind of datasets. Section 4 illustrates the methodology of the works. Section 5 contain the results of the studies and the last Section 6 shall present the conclusion of the study.

2. Literature Review

2.1. Social Media Analytics

Social media analytics provides a strategic approach on business decision making in creating value and gaining the competitive advantage (Bekmamedova and Shanks, 2014). It has been widely used in various field including consumer and marketing intelligent, supply chain and e-commerce. The social media analytics also able to influenced and improved the customer relationship, brand recognition and marketing, speed and scale, decreasing the price and flexibility (Bekmamedova and Shanks, 2014). Several steps were involved in social media analytic that are collecting data, extracting the data and analyzing the data. The process then proceed to presenting the user-generated data in order to support decision making and producing insight towards the organization (Holsapple *et al.*, 2014). The previous study state that there were three crucial stage involved in social media analytics that are capturing data, understanding data, and presenting data (Fan *et al.*, 2014). The capturing stage involve the collection of useful and relevant data from various social media platform such as Twitter, Blog and Facebook. Various kind of capturing technique is used in this stage such as news feed, Application Programs Interface (APIs) or by utilizing crawling or scrapping tools.

After the dataset is collected, the pre-processing stage shall transformed the dataset into valuable data for the organization by using opinion mining, sentiment analysis, topic modelling and trend analysis. The popular technique used in social media analytics include data mining, network analysis, natural language processing, text mining, marketing analytics, customer insight and customer engagement (Fan *et al.*, 2006)(Chamlertwat and Bhattarakosol, 2012). Based on Kurniawati et al findings that conducted on 40 social media analytics “successful story”, social media has improved marketing strategies, lead to a better customer services and customer engagement (Kurniawati *et al.*, 2013). Moreover social media analytics also helps to increase organization’s capability to improve business processes and product innovations.

2.2. Sentiment Analysis

Social media has emerged to be a platform that is used to represent user’s sentiment thus it has a big impact in increasing the requirement of data and text mining in the field of sentiment analysis. Many research attempt within the area of data mining or data analysis over the past years has been devoted to automatically classify textual content primarily that are not limited to subject matter, source and language. This classification able to correctly sort and manage the information. Within the last one and half decades, academia, research groups, public and service industries are working rigorously on sentiment analysis, also referred to as, opinion mining, in order to extract and examine public mood and perspectives.

According to the previous studies, the sentiment analysis has grew to be one of the most active study area in natural language processing (NLP) since early 2000 (Kaplan and Haenlein, 2010; Jimenez-marquez *et al.*, 2019). A sentiment analysis also consist of the sentiments content in user reviews, comment or opinion in social media whereby it focus on classifying the text based on polarity of the subjective text (Pang and Lee, 2004; Li and Dash, 2010). The polarity of the text can be either positive, negative or neutral polarity. Previous research has divide sentiment analysis into two task which are first, to detect which text sentence contain the sentiment and second, to determine the polarity and the strength of the sentiment (Pang and Lee, 2004).

Keshavarz and Abadeh (2017) stated that sentiment analysis is a field which deals with classifying and analyzing subjective people sentiments, opinion and emotion towards a product, organizations, individuals and so on that is shown or expressed in a text or sentences such as review, tweets, forums, blog and news. The process that are involve in sentiment analysis include data collection, text preparation, sentiment identification, sentiment classification and result (Amrani *et al.*, 2018). Text pre-processing and dataset cleaning is perform in the text preparation process while the identification and analyzing step of the sentiment words are done in the sentiment identification step. Last but not least, the classification of the sentiment is perform in order to obtain the result.

Numerous algorithm or technique are usually used for traditional text classification and this algorithm can be applied towards sentiment analysis. The algorithms or methods can be divided into two groups known as Machine Learning algorithm and Lexicon-based algorithm. Figure 1 display the sentiment analysis technique as discussed by (Medhat *et al.*, 2014).

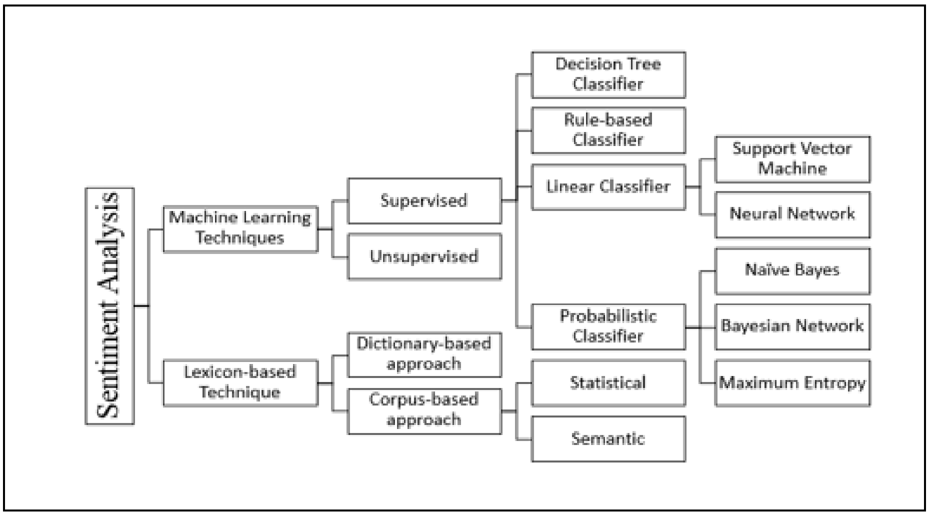

Figure 1. Sentiment Analysis classification techniques (Medhat *et al.*, 2014)

2.3. Machine Learning

Machine learning algorithm can be divided into two categories which are known as supervised and unsupervised learning. Supervised machine learning algorithm required pre-labelled dataset. The machine will try to understand a function from a labelled training dataset (Akkarapatty and Raj, 2016). Supervised machine learning algorithm is the more commonly used in sentiment analysis (Liu, 2012). In contrast, unsupervised machine learning algorithm do not required training dataset. Consequently, unsupervised techniques such as clustering and topic modelling are beneficial in case where the training dataset in unavailable or hard to be find.

Machine learning based sentiment analysis give vector value to a word using word embedding method. Then it will train the digitized sentences in the machine learning algorithm. In a traditional sentiment analysis, it is vital to know that the quality of the predefined labelled data will affect the analysis of the result gain in order to determine between positive or negative text (Do and Choi, 2015).

In this study, three supervised machine learning algorithm are used to build the classification models. The machine learning used are known as Naïve Bayes, Support Vector Machine (SVM) and Maximum Entropy. In this section, these algorithm were explain briefly.

3. Related Work

Within the Natural Language Processing, sentiment analysis is an active growing research area (Ravi and Ravi, 2015; Wang *et al.*, 2018). Thus, many research has been conducted related to sentiment analysis covering text dataset varies from Twitter, TripAdvisor, Facebook, e-mail, novels and others.

Hlee, Lee, and Koo (2018) reviewed and analyzed 55 research article related to the tourism and hospitality that was published in academic journals between January 2008 and December 2017. The author investigate the content-related characteristic of hospitality and tourism online reviews (HTORs) in different marketing segments and used heuristic-systematic model (HSM) to classify and summarize the characteristics that affect consumer's opinion in previous HTOR studies. They believed that their propose framework helps in the identification of research topic in extended HTORs literature and pointing out possible direction for future studies.

Xiang et al. (2017) offers a foundation for understanding the operational challenges and recognize several research path for social media analytics in hospitality and tourism area. They comparatively examined information quality related to online reviews on entire hotel population in Manhattan, New York using three major online review platforms which are TripAdvisor, Expedia and yelp through text analytics. The authors state that there exist an enormous inconsistencies in the representation of the hotel industry on these platforms. In addition, online reviews differ greatly in terms of their sentiment, semantic features, linguistic characteristic, rating and relationship between the features.

A study made an empirical analysis to discover the relationship between customer sentiment and online customer rating for hotel (Geetha *et al.*, 2017). The researchers considered using the lexicon of positive and negative words used by studies in (Hu and Liu, 2004). As for the classification algorithm, the researchers utilize Naïve Bayes classification technique to use the lexicon of word in order to match words from the document against the lexicon. Then, the classifier assigned the probability of the words being positive or negative. The result of the study find that there are consistency between customer rating and actual customer's feelings towards the hotel.

J. Bin Li and Yang (2017) implemented Chinese sentiment mining system towards TripAdvisor dataset as the evaluation sample. The proposed model was then compare with support vector machine and logistic regression models to compared their performance. As for the result, the proposed sentiment model outperform the two other models with accuracy of 81.3% and F-measure at 81.8%.
Chang, Ku, and Chen (2017) state that the real value of social media data is hardly recognize because of overloaded information and existing literature in analysing hotel reviews rarely provides decision making information and marketing insight to improve business services. The authors proposed an integrated framework consisting of several process to gain insight into hotel reviews and rating. The result revealed that the proposed approach outperform baseline algorithm with high precision and recall value which are 0.95 and 0.96.

Divyashree, L, and Majumdar (2017) used data mining and sentiment analysis techniques to analyse the polarity of the words from the reviews of all hotel on the TripAdvisor website. The studies use the data from Bengaluru Bangalore District Kartanaka Hotels dataset containing 6 attributes and 109 instances. The researcher used Partition around medoids (PAM) Algorithm to cluster the extracted data and utilizing J48 algorithm for the data classification which is then applied to the clustered dataset. As for the result, 86% of the words in the data was classified as positive words while 14% was classified as negative words.

Studies by Duan et al. (2016) were conducted using extensive and unique dataset of online user reviews for hotels across numerous reviews sites. Sentiment analysis was used in order to decompose the reviews into different dimensions to measure the service quality and performance of the hotel based on SERVPERF model. As the result, different dimensions of user reviews has significantly different effect in forming user evaluation and driving content generation.

Gao, Hao, and Fu (2015) has make an effort to discover and explore a set of three web services which are known as AlchemyAPI, Text2Data and Semantria. The authors then compare the sentiment analysis capabilities of these three services using reviews gain from TripAdvisor. They discover that these three web services has high accuracies since most of the reviews are classifies as positive.

4. Methodology

In order to implement the sentiment analysis towards the hotel reviews dataset in this study, several steps were followed. Figure 2 shows the operational framework for this study.

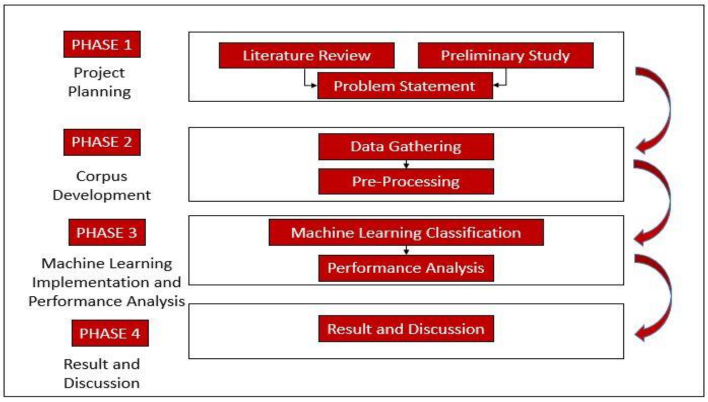

Figure 2. Operational Framework

4.1. Phase 1: Preliminary Study

Before the dataset was chosen from a specific hotel in Malaysia, a preliminary study was perform in order to find the best hotel in Malaysia. Interview techniques was used in order to gather information about Social Media Analytics towards hotel businesses. In this study, staff from sales and marketing departments in hotels was interview in order to gain information about their knowledge on social media websites for their hotel. This interview was also done to identify whether the hotel used social media marketing for their hotel. Based from the result gain from the interview session, it was known that most hoteliers has a good perception towards social media and social media analytics. However, the used of sentiment analysis was not implemented towards their hotel's reviews.

In order to identify which hotel was popular in Malaysia, Google Trend application was utilized. Five most popular hotel are selected to be compared using Google Trend application. The hotel are Grand Hyatt Kuala Lumpur, The Majestic Hotel Kuala Lumpur, Traders Hotel Kuala Lumpur, Shangri-La Hotel Kuala Lumpur, and Hilton Hotel Kuala Lumpur. Among these five hotels, it was discovered that The Majestic Hotel is frequently searched. Hence, the sentiment analysis will be perform on the reviews dataset of The Majestic Hotel available in the TripAdvisor website.

4.2. Phase 2: Corpus Development

Phase 2 of the studies was divided into two step which are data gathering steps and pre-processing steps. This two steps is crucial in order to prepare the data. Figure 3 shows the corpus development stage.

4.2.1 Data Gathering

Data gathering process was done by using a technique known as web scrapping to scrap the reviews of The Majestic Hotel in TripAdvisor page. The scrapping was done by using Python library known as Beautiful Soup that was specially designed for web scrapping HTML and XML files. As a result from the scrapping, the dataset consisted of 2434 rows of reviews data of The Majestic Hotel collected from 18th December 2012 until 17th November 2018.

4.2.2 Pre-processing

The reviews dataset undergoes several important pre-processing steps in order to prepare it for the machine learning implementation. The following steps are performed towards the reviews dataset.

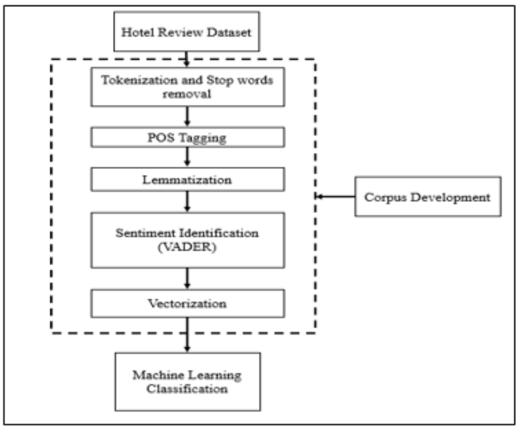

Figure 3. Corpus development stage

a) Tokenization and stop words removal. This steps is done in order to break the sentences in the reviews into terms, words and symbols or any other meaningful tokens by removing the punctuations marks (Ravi and Ravi, 2015). Tokenization is a steps that includes both data cleaning and feature extraction which resulted on output in the form of extracted list of words. Stop words are common words that are found in most text and it need to be remove as it does not contribute any meaning to the text.
b) Part of Speech (POS) Tagging (Akkarapatty and Raj, 2016) was used to allocate a part of speech to every word within the text data consisting of nouns, verbs, adjective and others. By using POS tag, the words that may convey user's opinion can be assign out and hence become very useful in sentiment analysis.
c) Lemmatization was performs to the dataset in order to maps the diverse forms of a words into it root form. In this study, the lemmatization is perform using WordNetLemmatizer provided by NLTK library which use the WordNet Database to lookup lemmas of words (Bird *et al.*, 2009).
d) The sentiment identification of the review dataset was done using Valence Aware Dictionary for Sentiment Reasoning (VADER) (Hutto and Gilbert, 2014). VADER integrate a lexicon and series of intensifiers, punctuation transformation, and emoticons along with some heuristic to compute the sentiment polarity of the text. The compound score that was generated using VADER was use to set the overall sentiment of the reviews. The review will be considered as positive if the compound score is more or equal to 0.5. The review will be considered as negative is the compound score is less than or equal to -0.5 and neutral if the score is between -0.5 and 0.5 (Karim and Das, 2018). This score can be seen in Figure 4.

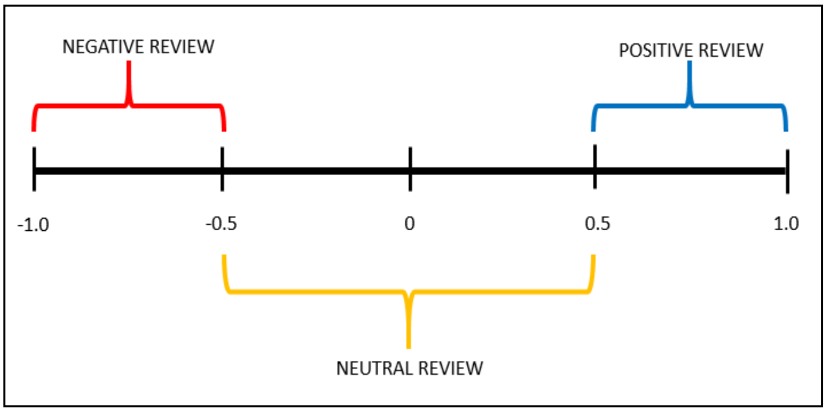

Figure 0. Overall sentiment classification based on VADER compound value

This study use CountVectorizer in order to produce a vectorise form of the data. It was also known as word embedding method (Wang *et al.*, 2018).

4.3. Phase 3: Machine Learning Implementation and Performance Analysis

This phase is divided into two step which are machine learning classification and performance analysis. The steps were done as follows. Figure 5 shows the overall flow of the overall process.

4.3.1 Machine Learning Implementation

Three type of machine learning classifiers were chosen to classify the sentiment in the dataset. The machine learning classifiers chosen are Naïve Bayes, Support Vector Machine and Maximum Entropy. Each classifiers was chosen based on their ability to classify the data. Naïve Bayes and Maximum Entropy classifiers were chosen because both techniques work with the text dataset that are represented as word counts while Support Vector Machine was chosen as it is a well-known powerful machine learning algorithm which work well on text classification.

Given a set of n document in vectors D = (d1, d2, …, dn), classified along a set C of m classes, C= {c1, c2, …, cm}, where m=3 where in this case are negative review, neutral review and positive review and n is the number of documents.

A. Naïve Bayes

Naïve Bayes (Bayes and Price, 1763) is known as a basic approach for developing classifiers models that assign class names to problem instances, spoke to as vectors of feature value, where the class labels are drawn from some limited set. Naïve Bayes classifier are based on Bayesian theorem and has been significantly employed within data mining research community. Naïve Bayes classifier has establishes accuracies similar to more advanced machine learning algorithms, while being less computationally complex (Su and Zhang, 2006). Naïve Bayes classifier is broadly used in text classification problem due to its effectiveness and simplicity [23] [24]. It work with text dataset that are represented as words count and used log linear decision rule where each independent parameter is assigned to each of class-word pair (Juan *et al.*, 2007). Bayesian classified the estimate probabilities of each classes of C for a given document D as:

$$P(c_m|d_n) = \frac{P(c_m)P(\vec{d}_n|c_m)}{P(\overrightarrow{d_n})} \quad (1)$$

where, in $P(\overrightarrow{d_n})$ is the probability that a randomly picked document has a vector $\overrightarrow{d_n}$ and $P(c_m)$ is the probability that a randomly picked document belongs to class c_m.

B. Support Vector Machine

Support Vector Machine (SVM) is one of the latest supervised learning techniques that had been developed (Vapnik and Cortes, 1995). SVM is mainly used for classification and regression and is widely used in many applications including the object detection and recognition, content-based image retrieval, text recognition, biometrics and many more. SVM classifies data by constructing an N-dimensional hyperplane in the future space that optimally divide the data into two categories (Joachims, 1998). The optimal linear discriminant function or classifier with the maximum margin give the best solution for the particular problem. SVM is chosen because of it is a well-known powerful machine learning algorithm (Wei *et al.*, 2012) and work well on text classification as it has the ability to generalize well in high dimensional feature space (Joachims, 1998). In addition, SVM eliminated the need for feature selection thus making the classification easier and it is also popular because of its robustness (Ranzato and Zanella, 2019). SVM is a method over a vector space where the problem is to find a best decision surface that separates the data into different classes. The decision surface is a hyperplane which ca be written as

$$wD + b = 0 \qquad (2)$$

Where, D is an arbitrary document to be classified, w and b are the constant that been set during the training. SVM is proposed to solve a linearly constrained quadratic programming problem so that the solution is always globally optimal

$$\min_{w} \frac{1}{2} \|w\|^2 + D \sum_{i}^{m} c_m \qquad (3)$$

with constraints

$$y_n(d_n w + b) \geq 1 - c_m | c_m \geq 0, \forall_m \qquad (4)$$

We consider a multi-class classification in this paper where we adopted One-Against-the-Rest approach. With this approach, k classes are defined as a collection of $\frac{k(k-1)}{2}$ of binary classification problems. The C classifier constructs a hyperplane between class m and the other k - 1 classes.

C. Maximum Entropy

Maximum Entropy (ME) is an approach of estimating the probability distribution that broadly used for a selection of natural language tasks, such as a part-of-speech tagging, text segmentation, and language modelling (Nigam *et al.*, 1999). This method has a unique solution which can be discovered by using the improved iterative scaling algorithm. Different with Naïve Bayes and SVM, the ME classifier do not assume that feature are independent towards each other. ME is use to estimate probability distribution from the data based on the principle of if there is no prior information or knowledge about the data, thus it should be randomly or uniformly distributed (Nigam *et al.*, 1999). ME model estimates chances based on the principle of creating as few assumptions as possible, apart from the limited imposed obtain from the training process (El-Halees, 2007). Similar to Naïve Bayes, ME also a popular technique for text classification and work with text [26].

The maximum entropy model estimates probabilities based on the principle of making assumptions other than constrained imposed (El-Halees, 2007). In this paper, a maximum entropy model assigns a class, C of each word in document, D. Conditional distributed $p(C|D)$ is computed(Berger *et al.*, 1996):

$$p(C|D) = \frac{1}{Z(D)} \exp(\sum_{i=1}^{n} \alpha_i f_i(D, C)) \qquad (5)$$

where Z(D) is a normalization function which is computed as:

$$z(D) = \sum_{c=1}^{m} \exp(\sum_{i=1}^{n} \alpha_i f_i(D, C)) \qquad (6)$$

where, αi must be learned by estimation by a iterative way using algorithms such as L-BFGS algorithm (Malouf, 2002). In this context, we also consider as a multi-class classification same as SVM.

4.3.2 Performance Evaluation

When using different type of classifiers, an assessment should be performed to analyses the classifiers performance when it is predicting the class labels. In order to analyses the performance of each of the classifiers, three measurement was used which known as accuracy, precision and recall value (Dumais *et al.*, 1998). Accuracy is the most intuitive performance measure and a high accuracy indicate to the best model. Accuracy measure how many text or data is correctly classified into their classes. Precision is a measure of the ability of a classifier to identify only the relevant data, while recall is a measure of the ability of a model to find all of the relevant cases within a dataset. A high value for precision and recall indicates that the classifier is performing well as it classifies most of the positive sample and does not classifies many sample that should not be classified. Their equation are as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where TP is denoted as True Positive value, TN is the True Negative value, FP is the False Positive value and FN is the False Negative value.

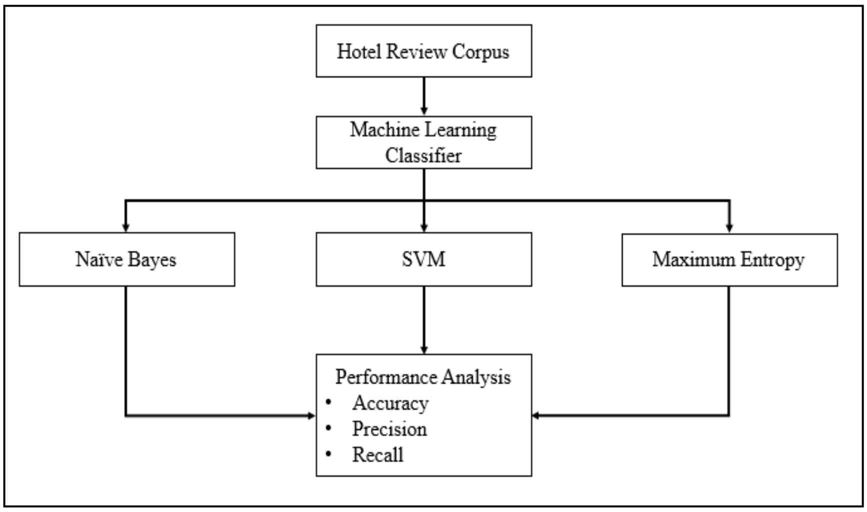

Figure 5. Process flow of overall process

4.4. Phase 4: Result and Discussion

The last phase of the study is the result and discussion phase. This will be discussed on the next section.

5. Experiment Result

The experiment were implemented by using Anaconda Navigator toolbox as the programming tools and Python as the programming language for the scrapping, pre-processing, modelling and performance analysis steps.

5.1. Data Exploration

In total, the number of the reviews gain from The Majestic Hotel was 2434 rows of text reviews were used for the sentiment analysis. The review dataset was stored in the form of data frame by using the pandas library in Python. Table 1 shows the example of the reviews dataset in the form of data frame.

Table 1. Example of review data set

REVIEW BODY	
0	This was our first stay at The Majestic, Kuala Lumpur. We had the straits room with the club access and butler. Food was fantastic. We had breakfast in the room every morning and had dinner at the Contango restaurant. The service was superb and the attention to detail was second to none. We will be back.
1	We are regular travelers to Kuala Lumpur and have stayed at many of the city's top hotels. For a change we chose the Majestic, a bit out of the center of the city but nevertheless within easy reach of all the attractions. The Majestic runs a regular scheduled shuttle bus between Sentral Station, Mid-Valley Mall, Bukit Bintang and Kuala Lumpur CC but you'll need a cab or Grab ride to get around. Be warned, Kuala Lumpur traffic is getting worse by the year but that's another story. We had a room in the newer tower (only 15 floors) but next trip we'll opt for a room in the old section. This refurbished colonial hotel is worth a stay, even if only for one time, to experience the colonial ambience and a level of service second to none. From the grand entrance to the Smoke House bar to the Orchid Conservatory to the indulgent bath with TV in the rooms, this grand hotel is a unique experience.
2	I was recommended to stay here during a recent 3 days 2 nights work trip and absolutely loved coming back here after work! Made me not want to leave my room at all with that bathtub and tv! The concierge team were also most helpful and obliging when they helped to hold the shuttle going to Sentral Station for me as I was a little late due to heavy traffic trying to rush back to the hotel on my last day. Pity I didn't have time to enjoy the breakfast although my husband has many times when he does his work trips over! Am looking forward to bringing family over soon and will be back.

The dataset undergoes the pre-processing steps in order to produce a clean dataset before it was given to the machine learning classifier. Figure 6 shows the word cloud that contained the clean reviews.

Figure 6. Word Cloud

We can see from Figure 6 show the most of the word are indeed related to the hotels such as rooms, staff, suite, and food. Some words are more related to the customer's experience with the hotel such as friendly, wonderful, disappointed and others. Table 2 shows the word frequency in the clean reviews dataset.

Table 2. Word Frequency

Word	Word Frequency
hotel	5404
room	4000
stay	2577
service	2017
majestic	1846
good	1619
breakfast	1371
great	1173
tea	1059
food	1028

As we can see, the word 'hotel' was mentioned 5404 times in the clean reviews dataset followed by 'room' with 4000 times and 'stay' for 2577 times.

5.2. Sentiment Identification

Sentiment identification of the reviews was done by using VADER. Three classes of the reviews was identified which are positive, negative and neutral reviews. Figure 7 shows the classes in the reviews.

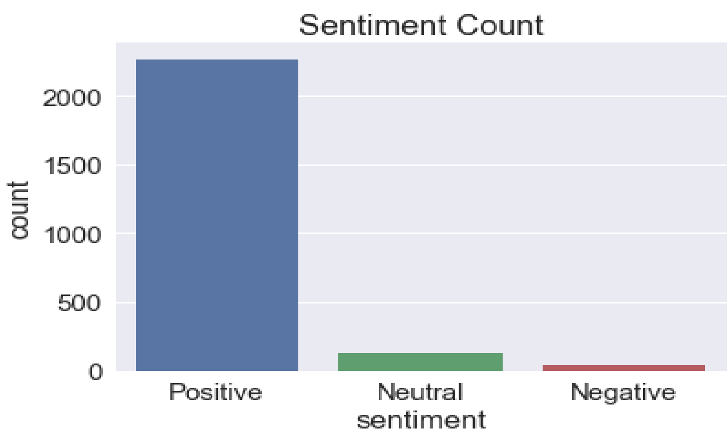

Figure 7. Number of sentiment classes in The Majestic Hotel review

There are 2268 reviews that are considered as positive reviews, 125 reviews are neutral reviews while only 41 of the reviews were considered as negative reviews.

As discussed previously, VADER give four type of score which are positive, negative, neutral and compound score. The overall sentiment for the reviews are classified by using the compound value given by VADER. Table 3 below shows the example of the reviews with their compound score and sentiment classification.

Table 3. The review and the sentiment classification

Sentiment Class	Reviews	Compound Score
Positive reviews	Amazing colonial style hotel. Rooms are new, clean and well designed. The greatest compliments to the staff which is very kind and ready to answer to every kind of questions, but biggest thanks goes to Kumar which is very kind, always smiling and ready to help you with whatever you need.	0.9836
Negative reviews	Bad service, bad public relation especially if you were to do events and overall it's a really bad hotel! The pool is small and the food at the restaurants are really bad, i am so disappointed with everything that they offered.	-0.9524
Neutral reviews	I just booked the club room in the Majestic wing. I was told by reservation that I do not have access to the spa pool unless I'm having a spa treatment. I find this absolutely absurd. I can immediately think of a few other 5-star properties that allow all guests to use the spa pool. I'm taking my business to the Mandarin Oriental.	0.0000

5.3. Performance Analysis

In this section, Naïve Bayes, SVM and Maximum Entropy was used in order to classify the review dataset. The review data was split into training and testing dataset by using train_test_split function provided by sckit learn model selection library. The dataset was split into 70% of training dataset and 30% of testing dataset.

The performance analysis was done using the value of micro-average of both precision and recall value(Sokolova and Lapalme, 2009). This is due to the multi-class classification setup where micro-average value is preferable due to class imbalance. Table 4 shows the overall result for accuracy, precision and recall value that was explained in the previous section.

Table 4. Performance Evaluation

Classifier	Accuracy (%)	Precision (Micro)	Recall (Micro)
Naïve Bayes	89.30	0.8930	0.8930
SVM	90.67	0.9067	0.9067
Maximum Entropy	92.16	0.9216	0.9216

Figure 8 shows the performance accuracy gain between three classifiers. As shown from the figure, Maximum Entropy gain the highest value of accuracy compared to the other two classifiers with the value of 92.16%. The second highest accuracy was obtain by SVM with 90.67% followed by Naïve Bayes with 89.30%.

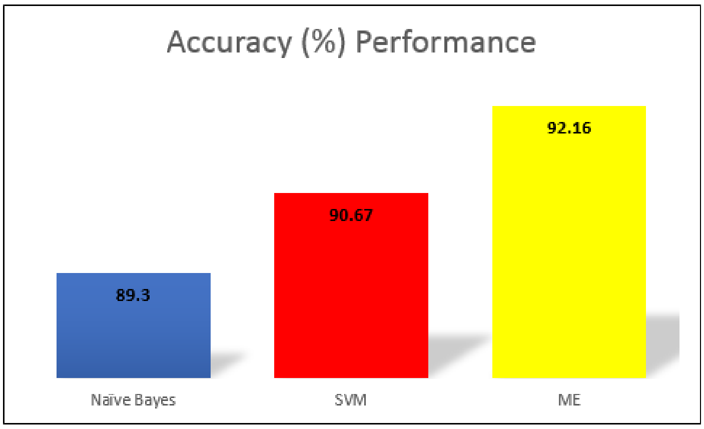

Figure 8. Performance evaluation using accuracy

Figure 9 shows the performance evaluation based on precision and recall value for each classifier. Note that this result is based on the micro-average value of each precision and recall as we are dealing with unbalanced dataset. As discussed in precious chapter, a high precision and recall value indicates that the classifier is performing well as it can classifies most of the positive sample. Based on Figure 9, Maximum Entropy gain highest value of micro-average precision and recall value which are 0.9216 respectively followed by SVM with 0.9067 and Naïve Bayes with 0.8930. Notice that the value for micro-average precision and recall is the equal. This was a correct behavior as every sample-class pair was given equal contribution to the overall metrics by the micro-average value. Based on the findings, Maximum Entropy has gain the highest accuracy and highest micro-average value of precision and recall which indicates that Maximum Entropy has outperform the other two classifiers.

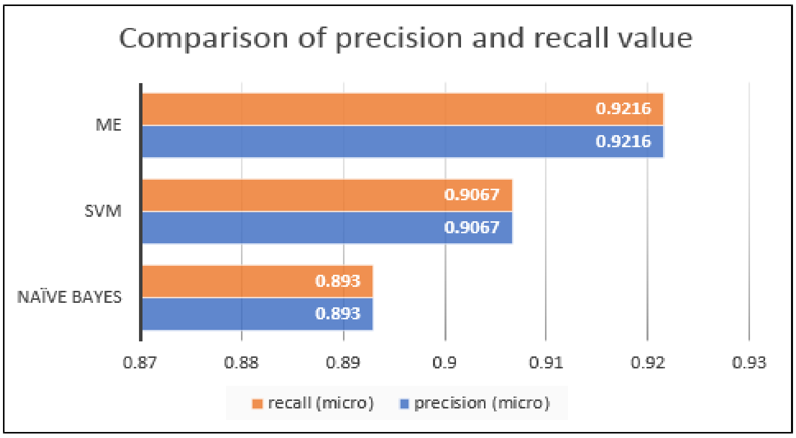

Figure 9. Performance evaluation on precision and recall value for each classifier

6. Conclusion

Many research attempt within the area of data mining or data analysis over the past years has been devoted to automatically classify textual content primarily based on subject matter, source, language and many more in order to be able to correctly sort and manage the information. Thus, various research related to sentiment analysis has been perform using different kind of text data.

The main goal of this study is to perform sentiment analysis towards hotel reviews by utilizing several classification techniques thus comparing the performance of each classifiers. The identification of the sentiment analysis study was determined through literature reviews that were related to sentiment analysis revealed that sentiment analysis has evolve to be one of the most active research area in Natural Language Processing. Various sentiment analysis has been conducted using different set of text dataset, however the number of sentiment analysis by using hotel reviews dataset is still lacking. Thus, this study aim to add additional sentiment analysis research on hotel reviews dataset.

The dataset undergoes pre-processing methods in order to develop the sentiment analysis model towards the review dataset. The pre-process include tokenization, stop word removal, sentiment identification, lemmatization and vectorization. The clean dataset was then fed to the classification model to be evaluated. The finding indicate that in the review, 2268 of the review was classified as positive reviews, 125 are neutral reviews and 41are classified as negative reviews. Among the three classifier used to classify the reviews, Maximum Entropy gain the highest accuracy at 92.16% followed by SVM and Naïve Bayes at 90.67% and 89.3%. Furthermore, Maximum Entropy also gain highest result for micro-average value of precision and recall at 0.9216 which outperform the other two classifier.

This study provides valuable implications on the sentiment analysis towards hotel reviews. In this project, by utilizing the compound value in VADER, it can lead to a better understanding of sentiment class and it make it easier for the machine learning classifier to train the model as the reviews has successfully being labelled. Machine learning classifier can perform well in the sentiment classification of the hotel reviews. From the business intelligence point of view, it was important to required insight about the product and services that a hotel provided. Thus, the hoteliers can utilized the social media in order to find what user think about them. For example, a hotel manager can extract the customer's reviews from their hotel website and use it for sentiment analysis of their hotel.

There are limitations that exist throughout the research. The first limitation is, the amount of dataset used in this project is low compared to other sentiment analysis research available. Next, the project only implement three type of machine learning classifier which are Naïve Bayes, Support Vector Machine and Maximum Entropy. The dataset used in this project was considered unbalanced because of the amount of positive reviews is greater compared to other classes of review thus effecting the accuracy of the classifiers. When it comes to the evaluation of the metrics, the research only used three typical measurement which are accuracy, precision and recall.

In the future, when it comes to the evaluation performance of the machine learning classifier, there are still many measures that can be utilized in order evaluate the result of this experiment. Future work for the sentiment analysis should also consider on how to tackle the unbalanced class in the review dataset to gain more accurate classification.

Acknowledgments

I wish to express my gratitude and appreciation towards Universiti Teknologi Malaysia for providing the access to the resources used in the completion of this research. I also would like to thank my supervisor, Dr Norziha binti Megat Mohd Zainuddin, for her encouragement, guidance, critics and motivation for me in making this project report a success.

References

A. I. Kabir, R. Karim, and S. Newaz, "The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R," Inform. Econ., vol. 22, no. April, 2018.

A. Juan, D. Vilar, and H. Ney, "Bridging the Gap between Naive Bayes and Maximum Entropy Text Classification," Pattern Recognit. Inf. Syst., vol. 01, no. September 2014, pp. 59–65, 2007.

A. L. Berger, S. A. Della Pietra, and V. J. Della Pietra, "A Maximum Entropy Approach to Natural Language Processing," Comput. Linguist., vol. 22, pp. 39–71, 1996.

A. M. El-Halees, "Arabic Text Classification Using Maximum Entropy," Islam. Univ. J. (Series Nat. Stud. Eng., vol. 15, no. 1, pp. 157–167, 2007.

A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities os social media," *Bus. Horiz.*, vol. 2010, no. 53(1), pp. 59–68, 2010.

B. Liu, Sentiment Analysis and Opinion Mining, no. May. 2012.

B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," 2004.

B. Y. W. Fan, M. D. Gordon, and W. M. Than, "The Power of Social Media Analytics," Commun. ACM, no. 57(6), pp. 74–81, 2014.

C. Holsapple, R. Pakath, and S. Hsiao, "Business Social Media Analytics : Definition , Benefits , and Challenges," Twent. Am. Conf. Inf. Syst. Savannah, pp. 1–12, 2014.

C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," Proc. Eighth Int. AAAI Conf. Weblogs Soc. Media, pp. 216–225, 2014.

E. K. Ikonomakis, S. Kotsiantis, and V. Tampakas, "Text Classification Using Machine Learning Techniques," WSEAS Trans. Comput., vol. 4, no. 8, pp. 966–974, 2005.

F. Ranzato and M. Zanella, "Robustness Verification of Support Vector Machines," 2019.

H. J. Do and H.-J. Choi, "Korean twitter emotion classification using automatically built emotion lexicons and fine-grained features," 29th Pacific Asia Conf. Lang. Inf. Comput. PACLIC 2015, pp. 142–150, 2015.

H. Keshavarz and M. S. Abadeh, "Knowle dge-Base d Systems ALGA : Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs," Knowledge-Based Syst., vol. 122, pp. 1–16, 2017.

J. Bin Li and L. B. Yang, “A Rule-Based Chinese Sentiment Mining System with Self-Expanding Dictionary - Taking TripAdvisor as an Example,” IEEE Comput. Soc., pp. 238–242, 2017.
J. Chen, Z. Dai, J. Duan, H. Matzinger, and I. Popescu, “Naive Bayes with Correlation Factor for Text Classification Problem,” 2019.
J. L. Jimenez-marquez, I. Gonzalez-carrasco, J. L. Lopez-cuadrado, and B. Ruiz-mezcua, “International Journal of Information Management Towards a big data framework for analyzing social media content,” *Int. J. Inf. Manage.*, vol. 44, no. September 2018, pp. 1–12, 2019.
J. Su and H. Zhang, “Full Bayesian network classifiers,” Proc. 23rd Int. Conf. Mach. Learn., pp. 897–904, 2006.
Kurniawati, B. Nargiza, and S. Graeme, “The Business Impact Of Social Media Analytics,” ECIS 2013 - Proc. 21st Eur. Conf. Inf. Syst., vol. 48, 2013.
K. Nigam, J. Lafferty, and A. McCallum, “Using Maximum Entropy for Text Classification,” IJCAI-99 Work. Mach. Learn. Inf. Filter., vol. 1, pp. 61–67, 1999.
K. Ravi and V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, vol. 89, no. November. 2015.
L. Wei, B. Wei, and B. Wang, “Text Classification Using Support Vector Machine with Mixture of Kernel,” J. Softw. Eng. Appl., vol. 05, no. 12, pp. 55–58, 2012.
M. Geetha, P. Singha, and S. Sinha, “Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis,” Tour. Manag., pp. 43–54, 2017.
M. Hu and B. Liu, “Mining and Summarizing Customer Reviews,” in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.
M. Karim and S. Das, “Sentiment Analysis on Textual Reviews,” IOP Conf. Ser. Mater. Sci. Eng., vol. 396, no. 1, 2018.
M. Mariani, G. Di Fatta, and M. Di Felice, “Understanding Customer Satisfaction with Services by Leveraging Big Data: The Role of Services Attributes and Consumers’ Cultural Background,” IEEE Access, vol. 7, pp. 8195–8208, 2019.
M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Inf. Process. Manag., vol. 45, no. 4, pp. 427–437, 2009.
N. Akkarapatty and N. S. Raj, “A Machine Learning approach for classification of sentence polarity,” 3rd Int. Conf. Signal Process. Integr. Networks, SPIN 2016, pp. 316–321, 2016.
N. Bekmamedova and G. Shanks, “Social media analytics and business value: A theoretical framework and case study,” Proc. Annu. Hawaii Int. Conf. Syst. Sci., pp. 3728–3737, 2014.
N. Divyashree, S. K. K. L, and J. Majumdar, “Opinion Mining and Sentimental Analysis of TripAdvisor.in for Hotel Reviews,” Int. Res. J. Eng. Technol., vol. 4, no. 11, pp. 1462–1467, 2017.
N. Li and D. Dash, “Using text mining and sentiment analysis for online forums hotspot detection and forecast,” Decis. Support Syst., vol. 48, no. 2, pp. 354–368, 2010.
R. Malouf, “A comparison of algorithms for maximum entropy parameter estimation,” in Proceedings of the Sixth Conference on Natural Language Learning, 2002, pp. 49–55.
S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python : Analyzing Text with the Natural Language Toolkit. 2009.

S. Dumais, J. Platt, D. Heckerman, and M. Sahami, “Inductive learning algorithms and representations for text categorization,” in 7th International Conference on Information and Knowledge Management, 1998.
S. Hlee, H. Lee, and C. Koo, “Hospitality and tourism online review research: A systematic analysis and heuristic-systematic model,” Sustain., vol. 10, no. 4, 2018.
S. Gao, J. Hao, and Y. Fu, “The Application and Comparison of Web Services for Sentiment Analysis in Tourism,” in *12th International Conference on Service Systems and Service Management (ICSSSM)*, 2015.
T. Bayes and R. Price, “An essay towards solving a Problem in the Doctrine of Chances,” Philos. Trans., pp. 370–418, 1763.
T. Joachims, “Text Categorization with Support Vector Machines : Learning with Many Releveant Features,” Eur. Conf. Mach. Learn. (ECML). Chemnitz, Ger., pp. 137–142, 1998.
V. Vapnik and C. Cortes, “Support-Vector Networks SVM.pdf,” vol. 297, pp. 273–297, 1995.
W. Chamlertwat and P. Bhattarakosol, “Discovering Consumer Insight from Twitter via Sentiment Analysis.,” J. Ucs, vol. 18, no. 8, pp. 973–992, 2012.
W. Duan, Y. Yu, Q. Cao, and S. Levy, “Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach,” Cornell Hosp. Q., vol. 57, no. 3, pp. 282–296, 2016.
W. Fan, L. Wallace, S. Rich, and Z. Zhang, “Tapping The Power of Text Mining,” Commun. ACM, vol. 49, no. 9, 2006.
W. He, H. Wu, G. Yan, V. Akula, and J. Shen, “Information & Management A novel social media competitive analytics framework with sentiment benchmarks,” vol. 52, pp. 801–812, 2015.
W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014.
Y. A. L. Amrani, M. Lazaar, K. Eddine, and E. L. Kadiri, “ScienceDirect ScienceDirect Random Forest and Support based Hybrid on Vector Intelligent Machine Approach to Sentiment Analysis Random Forest and Support Vector Machine based Hybrid Approach to Sentiment,” Procedia Comput. Sci., vol. 127, pp. 511–520, 2018
Y. Chang, C. Ku, and C. Chen, “Social media analytics : Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor,” *Int. J. Inf. Manage.*, no. September, pp. 0–1, 2017.
Y. Wang, L. A. Kung, W. Y. C. Wang, and C. G. Cegielski, “An integrated big data analytics-enabled transformation model: Application to health care,” Inf. Manag., vol. 55, no. 1, pp. 64–79, 2018.
Z. Xiang, Q. Du, Y. Ma, and W. Fan, “A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism,” Tour. Manag., vol. 58, pp. 51–65, 2017.