TEXT MINING FOR PUBLIC PERCEPTION ANALYSIS IN THE TELECOMMUNICATION COMPANIES IN RWANDA: CASE OF MTN RWANDA

Joel Manzi Murwanashyaka & Dr. Annie Uwimana, PhD

The Strategic JOURNAL of Business & Change MANAGEMENT
ISSN: 2312-9492 www.strategicjournals.com
Vol. 10, Iss.4, pp 1167 – 1174. November 14, 2023. www.strategicjournals.com, ©Strategic Journals

TEXT MINING FOR PUBLIC PERCEPTION ANALYSIS IN THE TELECOMMUNICATION COMPANIES IN RWANDA: CASE OF MTN RWANDA

[1] Murwanashyaka, J. M., & [2] Uwimana, A.

[1] Master of Data Science (Actuarial Science) Student - College of Business Studies, University of Rwanda
[2] Senior Lecturer - College of Business Studies, University of Rwanda

Accepted: October 30, 2023

DOI: http://dx.doi.org/10.61426/sjbcm.v10i4.2811

ABSTRACT
The utilization of social media platforms has become indispensable for a significant number of individuals as a means to articulate their viewpoints and emotions. Organizations are increasingly recognizing social media platforms as a crucial instrument for obtaining pertinent information pertaining to various subjects and policies. Social media has had a significant rise in global usage, with a multitude of individuals utilizing these platforms to express their viewpoints on various aspects of social life, legislation, and politics. Twitter is a widely utilized social media platform for individuals to voice their viewpoints. The significant rise in information dissemination on Twitter throughout the year has positioned it as a prominent data source and a preferred option for conducting research on customer opinions and sentiment analysis. MTN Rwanda Plc has employed the use of questionnaires as a means of conducting customer opinion surveys for about 24 years since its establishment in Rwanda. This study aligned with the MTN project aimed at facilitating MTN Rwanda's access to consumer sentiments regarding newly announced products and services. It achieves this by analyzing the polarity of customer tweets on MTN Rwanda's Twitter accounts. The study examined the Twitter tweets of MTN Rwanda with the aim of doing sentiment analysis. The objective was to extract subjective sentiments, emotions, and public audience thoughts and opinions. The findings served as input to the MTN Rwanda customer experience department. The researcher utilized natural language processing and text mining methodologies to analyze and evaluate the public perceptions and expectations of MTN Rwanda's Twitter popularity rate. This involved employing text mining and analytics techniques. The findings indicated that there is a prevailing positive public impression and sentiment towards the tweets of MTN Rwanda. Moreover, the artificial neural network classifier predicts that this good sentiment is likely to persist in the future. The practical applications of the study were also discussed. It was expected that the study findings would benefit bilingual business entities on a global scale.

Key Words: *Text Mining; Text Analytics; Social Media; Sentiment Analysis*

CITATION: Murwanashyaka, J. M., & Uwimana, A. (2023). Text mining for public perception analysis in the telecommunication companies in Rwanda: case of MTN Rwanda. *The Strategic Journal of Business & Change Management,* 10 (4), 1167 – 1174. http://dx.doi.org/10.61426/sjbcm.v10i4.2811

INTRODUCTION

Social media platforms facilitate rapid electronic communication of various content types, encompassing personal details, files, videos, and images. The utilization of social media, exemplified by platforms like Facebook, Instagram, and Twitter, has experienced a swift global upsurge (Ortiz-Ospina, 2019). Presently, around one-third of the world's populace engages with social media (Zahoor & Qureshi, 2017). This surge in social media users has prompted businesses to capitalize on these platforms for marketing endeavors and customer engagement, thereby establishing social media as a potent marketing instrument (Zahoor and Qureshi, 2017; Ashley and Tuten, 2015; Oliverio, 2018; Hanaysha, 2022).

Businesses leverage social media to comprehend and chart their customers' actions through online information searches, reviews, and promotional activities (Hanaysha, 2022; Zeng & Gerritsen, 2014). The advantages of social media marketing for enterprises encompass amplified sales, heightened online platform traffic, enhanced brand recognition, improved brand perception, reduced marketing expenses, and augmented user engagement on platforms via incentives for content sharing (Felix et al., 2017).

Telecommunication companies across the globe have been active in social media, such as Twitter, to communicate and promote their deals to their followers and the public in general. As people in various nations use more than one language for communication, airlines' Twitter handles also tend to interact with their followers using multiple languages. For instance, MTN in Rwanda uses Kinyarwanda and French. The use of a specific language intends to target a group of customers who can easily understand it to maximize engagement. MTN practically every large corporation has a Twitter account to keep track of client comments on their services or products. Twitter is a powerful microblogging platform that allows users to post status updates (called "tweets"). These tweets contain a lot of human expressions, such as likes, dislikes, and contributions to many issues. However, it is not known whether the use of one language over the other tends to create more engagement on social media platforms.

As information technology is also evolving the data management is evolving. Data has grown above the level of databases and data warehouses with the increase of data on the web and social media. And this has brought the need of data mining, which is the process of finding anomalies, patterns, and correlations within large data sets to predict outcomes. The future trends and behaviors which allow the businesses to make proactive, knowledge-driven decisions are predicted using the data mining tools. The prospective analyses offered by data mining is automated and move beyond the analyses of past events provided by the retrospective tools of decision support systems. Data mining tools help us to answer the business questions that were traditionally time-consuming to resolve. Most companies collect and refine massive quantities of data forehand.

Research Problem

Rwanda's mobile telecommunication services sector has been experiencing challenges of customer complains towards network availability and poor service and customer have been expressing their opinions about their experience towards telecom services and products in Rwanda.

The telecommunication companies have engaged in increasing their network coverage in Rwanda and converting their 3G technology to 4G technology as means of trying to address customer complains. Customers put their sentiments on twitter frequently and more influential than other social media because they are made public for all to see. MTN Rwanda is the predominant player telecommunication company in Rwanda, with market share above 65% and Value share above 70% as per data published by Rwanda Utility and Regulatory Authority in quarter 1 of 2022 (RURA, 2022). This motivated me to develop this paper, to develop a model that could predict Public Perception towards the MTN Rwanda services and products using twitter data.

The key questions to be answered in this study were;

- How does the public feel about MTN Rwanda's service and product from April 2019 to April 2022.
- What is the best machine learning technique should be applied to predict the sentiment polarity behind MTN's tweets?

METHODOLOGY

The study extracted data from MTN Rwanda Twitter account, Twitter is a social networking and microblogging platform which allows users to tweet real time messages. These tweets are restricted to 140 characters in length each and which most of time leads to users to use abbreviations and emoticons. The data set used consists of 72,000 tweets and 62,000 retweets that have been extracted using the Twitter Application Programming Interfaces (API) and were extracted using the keyword (@MTNRwanda).

The variables which are in the data set are: The polarity of the tweet (positive or negative), which is the target variable; the date of the tweet; unique ID of the tweet; user referring to the username of the tweeted text. The study limited the data collection to only tweets and retweets which are in English language, other languages such Kinyarwanda and French used in Rwanda. The study excluded them from my dataset since they have no sentiment associated dictionaries for analysis. The study used the data extracted from April 2019 to April 2022. The R software was used in data mining, processing and analysis.

Machine Learning Algorithms for Sentiment Prediction

The study used the machine learning techniques to predict the public sentiments behind the MTN Rwanda's tweets. The existing set of algorithms have been developed to enable computers to learn such target mapping functions from data and make accurate predictions.

The selection of the algorithms to be used is depending on different criteria such as the size, structure and data type, distribution of data entries, complexity, latency and speed of the model itself. The employed models are detailed below:

Logistic regression

This is a machine learning model which was borrowed from the field of statistics. It is used for classification problems and model the occurrence probability of a certain event or existing event.

In my case, I have the data set called M = $\{(x1, y1)$; … ; $(xn, yn)\}$ with

$xii \in \mathbb{R}n{*}p$ and$yii \in \{Pos, Neg\}$; where $n \gg$, the Logistic Regression will model the occurrence probability that an input χi belongs to either positive (Pos) or negative (Neg) sentiment classes. Formally, we can write this as:

$(xii) = (Y = Pos\ or\ Neg \mid xii)$ (11)

where, i = 1,2,3, … , n, with n representing the number of data observations.

Naïve Bayes

The Naïve Bayes classifier is a probabilistic machine learning algorithm which is basing on Bayes theorem, and it is used in wide variety of classification tasks. It supports strong independence and equality assumptions between the variables.

The Bayes' theorem finds the probability of an event to occur given the probability of another event that has already occurred. it is expressed as follows:

$(A \mid B) = P(B \mid A)P(A)P(B)$ (12)

where, A and B are the events, P (A) is the occurrence probability of event A and P(B) is the occurrence probability of event B.

By the use of the above theorem, I want to find the probability of event A given that event B has happened.

Finally, P (A|B) is the probability of event A after evidence happens (a posteriori probability of B). Then, in relation to our data set and sentiment classification task, we applied the Bayes' theorem as follows:

$(y|X) = P(X|y)P(y)P(X)$ (13)

where, y is class variables (positive and negative) and X is set of dependent feature vector of size n, where: X = $(x1, x2, x3, \ldots, xn)$ and P (X|y) means the probability of occurrence of sentiment class given any MTN Rwanda' tweets.

RESULTS AND INTERPRETATION

Table 1: Detailed characteristics of data sets used.

Description	Count
Number of tweets	3,254
Number of retweets	147
Tweets without retweets	3,107
Number of Rwandan Twitter users (As of 2022)	**162,300**
MTN Rwanda Followers	110,834
MTN Rwanda Friends	1,340
Tweets favorites (Total Popular/liked tweets)	10,971
Tweet listed (Subscribers as interested in MTN Rwanda Posts)	89
MTN Rwanda tweet statuses (Authenticated Users tweeted)	69,944

Source: *https://datareportal.com/reports/digital-2022-rwanda and calculations from MTN Rwanda Twitter Account.*

By applying the natural language Emolex dictionary based on English language, eight basic emotions are: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust were computed.

Under the period of the study, the majority Twitter users expressed themselves positively with 58.09 % score, second being neutral with 31.9 % score, and last being negative sentiment with 10.02 %. The findings may prove that the reactions of the public towards the MTN Rwanda tweets on its services, products and communications are predominantly positive, which shows the trust to the MTN Rwanda's services and products. This also gives an assignment to look after those who expressed negative expression and those who held their feelings to be neutral so that going forwards, they should be in category of expressing positivity to MTN Rwanda's services and products. Figure 1 showed tabled polarity in a graph.

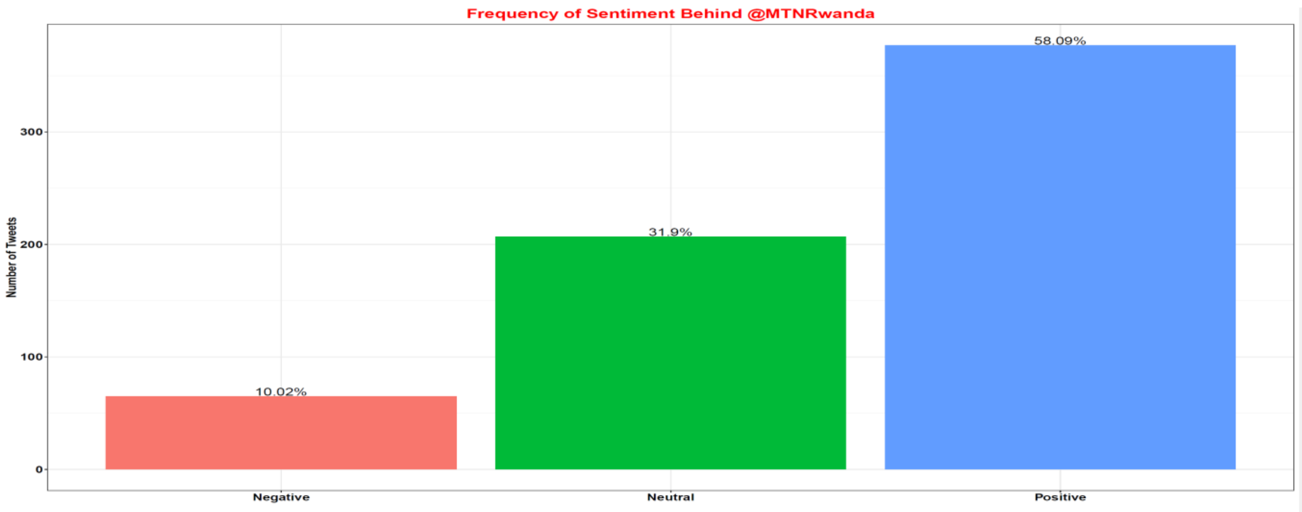

Figure 1 Public sentiment classes (Source: Author's plot.

The study used a time series sentiment analysis to understand the trends and patterns of how the MTN Customers responded to MTN Rwanda tweets over time. As stated in the introduction, the objective (2) was to understand the public sentiment towards the MTN Rwanda tweets. To accomplish this goal, the study created two visual displays for Twitter sentiment trends and the comparison word cloud. The sentiment trends revealed the changes in public sentiments over time, which might correspond to specific events, new products launched, new services launched and new campaigns. The comparison word cloud was found to be a powerful tool to understand the discussion or interests of the public in the period under review.

Figure 2 presents MTN Rwanda Twitter sentiment trends, displaying the percentages of positive and negative tweets per month, respectively, for the research period of April 2019 to April 2022. The spikes and vales in these trends reveal how the public sentiment changed in respect with the MTN Rwanda campaigns launched, new products launched, and services launched, especially on specific # hashtags.

By analyzing the trend, the study noticed a hike in positive sentiments in January of 2022 which corresponds to the #BivaMoMotima campaign which was a promotion for Mobile Money users to send and receive money and get a chance to win prizes among them was a new brand car. These observations reveal how the public sentiments can be determined by events, and that the MTN Rwanda should take into consideration the social media responses when making launching of its products and services.

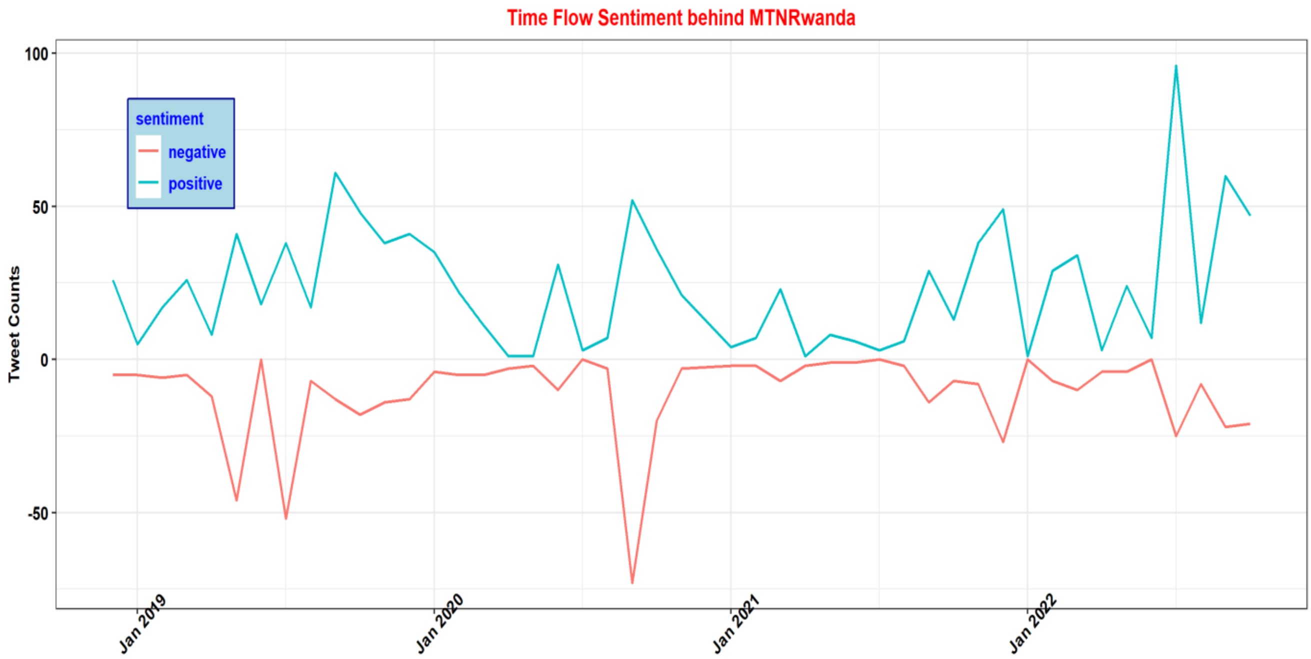

Figure 2: Public sentiment trend over time.

Predictive Machine Learning Algorithm Results

The different five machine learning models were trained on the dataset to optimally classify the sentiment classes prevailed in the utilized data set. The several performance metrics such accuracy, F-measure, Cohen's Kappa statistics, area under the curve were recorded for each trained model. The table below shows clearly the performance of each individual algorithm in terms of performance metrics. However, the optimal (outperformed) models were selected basing on both F-measure and cohen's kappa statistics with reference to the several machine learning studies conducted on binary classification problems (Wirtz, 2007). In essence, the cohen's kappa statistic is mostly used for binary classification problems in machine learning and used to measure of how closely the instances classified by the *machine learning classifier* matched the data labeled as *ground truth* (Mousumi Banerjee, 2008). Also, it is controlling the accuracy of a random classifier as measured by the expected accuracy (Mousumi Banerjee, 2008) . At the other hand, F-measure or F1 score which is defined as the harmonic mean of precision and recall combines classifier's precision and recall computing the weighted average performance rate of trained binary classifier (Gaussier, 2005). In this study, we have used also F measure to select the optimal model by referring to the other literature reviewed in the field of machine learning classification problems. George and colleagues agreed that F1 is usually more useful than accuracy, especially if you have an uneven class distribution (George Hripcsak, 2005) as we do in our case.

Therefore, basing on both cohen's kappa and F-measure performance metrics, the selected outperformed model will be used to predict the public sentiment behind the MTN Rwanda communication on Twitter platform. The model results are displayed in the below table:

Table 2: Performance measures of machine learning algorithms.

S/N	ML Model	Accuracy	F-Measure	Cohen's Kappa	Area Under ROC
1	Logistic Model	0.8102631	0.8847448	0.609672	0.8259019
2	CART	0.8126050	0.7614916	0.714392	0.8323954
3	Random Forest	0.8146714	0.7884565	0.630834	0.8300851
4	**Adaptive Boosting Machine**	**0.9172889**	**0.8922437**	**0.898686**	**0.9384063**
5	Neural Network	0.8166001	0.8735187	0.762014	0.8381117

From the above table of predictive outputs, the Logistic Regression's accuracy is 81.02% with an F-measure (F1 score) of 88.4%, Cohen's Kappa statistics of 60.9%, and area under AUR curve of 82.6%. The CART showed an accuracy of 81.3% in prediction of the sentiment polarity of the test observations, with an F-measure of 76.14%, Cohen's Kappa statistics of 71.4%, and AUC of 83.2%.

The Random Forest algorithm attained a relatively higher accuracy of 81.4% compared to CART with the precision-recall proportion of 78.8%; Cohen's Kappa of 63.1% and relatively higher AUC of 83.0%, whereas Artificial Neural Network with an accuracy of 81.7% and an F-measure of 89.3% and finally, the Adaptive Boosting Machine is being considered as winner and outperformed classifier basing on its higher accuracy, Fmeasure, Cohen's kappa and area under ROC compared to other trained models.

Among the five-trained classifiers, the greatest results have been achieved by the Adaptive Boosting Machine (AdaBoost) algorithm on all measures having the highest accuracy of 91.7%, and F-measure of 89.2%, a Cohen's Kappa statistics of 89.8% that suggests the perfect agreement between actual and predicted sentiment polarities. The AdaBoost is also having the greater discriminating power of positive from negative sentiments since it covers 93.8% of the area under the ROC curve, as shown in Table above. Thus, the identification of optimal out-performed model is crucial in machine learning, especially when it comes to the effective and accurate prediction in classification problems.

The literature tailed above gives us the right decision to select Adaboost classifier as the superior optimal performing model relatively to others due to its highest weighted average recall and precision (Fmeasure) and Cohen's kappa**.** Therefore, such empirical results suggest that the Adaboost should be used in predicting public sentiment polarities towards MTNRwanda communication in near future.

DISCUSSIONS

Public feel about MTN Rwanda's service and product over a given period.

From the analysis the study conducted to MTN Rwanda tweets, the researcher noted that 58.09% of sentiments towards MTN Rwanda tweets are positive which indicates that the majority are happy of MTN communications on its new products and services. Also, the researcher noted neutral sentiments to be above negative reactions. Neutral sentiments are at 31.9 % while negative sentiments are at 10.02 %.

In the time series analysis conducted the researcher Noticed that in the period of June 2020, the negative sentiments was higher than positive sentiments and this was due to bad network experience during Covid-19 lockdown when people were complaining about bad network since all of them were working from home and their internet was very bad also by analyzing the trend, I noticed a hike in positive sentiments in January 2022 which corresponds to the #BivaMoMotima campaign which was a promotion for Mobile Money users to send and receive money and get a chance to win prizes among them was a new brand car. These observations reveal how the public sentiments can be determined by events, and that the MTN Rwanda should take into consideration the social media responses when launching its products and services.

Best machine learning technique to be applied to predict the sentiment polarity behind MTN's tweets

Five machine learning models namely, Logistic Model, CART, Random Forest, Adaptive Boosting Machine and Neural Network are compared and identified Adaptive Boosting Machine as the efficient model with good accuracy, this was achieved by use of scoring metrics which are Accuracy, F-Measure, Cohen's Kappa and Area Under ROC and Adaptive Boosting Machine learning outperformed others. The comparison of the five algorithms the study concluded that Adaptive Boosting Machine Learning is better over the four algorithms. Hence Adaptive Boosting Machine Learning technique was applied to predict for the future of public sentiments towards MTN Rwanda tweets.

CONCLUSION AND RECOMMENDATION

In this research, the study conducted a twitter sentiment analysis in the context of public feedback to the MTN Rwanda Twitter posts. Two techniques were used to perform text classification and prediction which are lexicon-based approach and machine learning – based approach respectively.

In this research, it also denotes how sentiment analysis results are useful to identify trends and patterns of public sentiment driven by unique communications about MTN Rwanda Products and services. The research's findings have two implications. First, the public sentiment towards MTN Rwanda can be influenced by its products and services communicated to its twitter handler, so it is of paramount importance for MTN Rwanda to consider public's opinions expressed via social media in the process of new products and services evaluation. Second, sentiment analysis has demonstrated to be an important tool for both recognizing current sentiments and predicting future sentiments.

Social media sentiment analysis using novel machine learning and lexicon-based approaches should be integrated into MTN Rwanda strategy to mine the public's sentiment, in effort to hear their voice, promote public trust, accountability, and increase transparency of the MTN Rwanda's activities.

There were some limitations to this study. First, being the twitter APIs access, the researcher was not able to give full access, and my access was limited for a period from 2019 yet the researcher wanted to extract data from 2011. Also, there was a limitation on language since Natural Language Processing dictionaries do not interpret Kinyarwanda, yet some tweets were in Kinyarwanda.

The study recommended that MTN Rwanda to start doing analysis on how its customers react about its services and products using its social media platforms since this is where a customer can freely express his/her sentiments.

REFERENCES

Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology & Marketing*, 32(1), 15-27.

Felix, R., Rauschnabel, P. A., & Hinsch, C. (2017). Elements of strategic social media marketing: A holistic framework. *Journal of Business Research*, 70, 118-126.

Gaussier, É. (2005). Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the Tenth International Conference on Information and Knowledge Management (pp. 168-169).

George Hripcsak. (2005). Evaluating Natural Language Processing Applications. *Journal of the American Medical Informatics Association*, 12(5), 509-510.

Hanaysha, J. (2022). The role of social media marketing in business performance: A conceptual framework and research agenda. *Journal of Marketing Communications*, 28(1), 69-87.

Mousumi Banerjee. (2008). Diagnostic tests (2nd ed.). CRC Press.

Oliverio, D. (2018). Social media marketing: A literature review and implications. Psychology & Marketing, 35(3), 261-271.

Ortiz-Ospina, E. (2019). Social Media Usage Worldwide. Our World in Data. https://ourworldindata.org/social-media

Wirtz, R. A. (2007). The role of engagement in customer loyalty and retention. *Journal of Retailing and Consumer Services*, 14(3), 271-280.

Zahoor, N., & Qureshi, I. M. (2017). Social Media as Marketing Tool: A Literature Review. *Journal of Marketing Management and Consumer Behavior*, 2(3), 23-28.

Zeng, F., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. Tourism Management Perspectives, 10, 27-36.