THE ROLE OF TRAVELER'S FEEDBACK IN TOURISM DEVELOPMENT: ANALYSIS OF TRAVEL REVIEWS ON THE PHILIPPINE TOURISM EXPERIENCE

Marjorie QUIAOIT[1] and Hideki FURUYA[2]

[1]Master's Student, Graduate School of International Tourism Development, Toyo University
(5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan)
E-mail: marj.quiaoit@gmail.com
[2]Member of JSCE, Professor, Department of Tourism, Toyo University
(5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan)
E-mail: furuya@toyo.jp (Corresponding Author)

Travel-related websites are considered an important part of tourism nowadays since they not only provide non-biased reviews of a destination or establishment for potential tourists, they are also used by tourism managers to understand the needs, wants and expectations of the market to further improve their offerings. However, a very limited number of researches have been conducted with regards to the latter. This research analyzes traveler's comments and feedback about their Philippine trip to gauge the country's performance as a tourism destination. A total of 1,717 travel reviews and blogs were gathered from three travel-related websites and analyzed using machine learning techniques such as Sentiment Analysis, Word2Vec Analysis, and Clustering Analysis. The findings revealed that reviews written about the Philippines are mostly positive, but it also points out the tourism components that tourists did not like and must be improved. Furthermore, proposed applications of the Word2Vec output for tourism promotion and development are presented and implications of the research findings in Philippine tourism policies are discussed.

Key Words : *travel review, traveler feedback, machine learning, tourism development*

1. INTRODUCTION

It is undeniable that the Internet now plays an important role in the tourism industry and became an indispensable part of travel. It started to gain ground in the mid-1990s when online travel agencies (OTAs) were first established and started a booking-oriented wave that resulted in an increase in online travel bookings from 19% in 2000 to 57% by 2008[1)]. However, the advent of Web 2.0 gave birth to a second wave of travel information websites called Travel 2.0[2)]. Similar to Web 2.0, the term "Travel 2.0" coined by Phocuswright refers to the interactive information sharing and collaboration of travel knowledge to co-create user-generated content (UGC) in a virtual community[3)].

Several tourism researchers studied how these UGCs affect the way people travel and found out that Travel 2.0 has greatly influenced tourist behavior. According to the UNWTO, consumers usually spend a large amount of time searching online and visit an average of 14 different travel-related sites, with three visits per site, before making an online purchase[4)]. Browning et.al[5)] and Chong et.al[6)] found out that online tourism has a strong influence on tourists deciding their next holiday destination while O'Connor[7)] noted that positive and negative reviews of a hotel provide insights that allow customers to evaluate their alternatives. Online reviews are found to be trustworthy, authentic, and useful[8)], thus more and more potential tourists rely on them before making travel decisions. Because of this, travel review sites (e.g. TripAdvisor and Lonely Planet), travel blog platforms (e.g. Travelblog and Off Exploring), and social media (e.g. Facebook and Instagram) are now considered the word-of-mouth of the digital age[9)].

Tourism, being a service-oriented industry, is heavily reliant on positive feedback and high customer satisfaction[10)] and the advent of Travel 2.0 where tourists themselves share their comments in real-time and are readily available to anyone with internet access poses great potential for the tourism industry. On one hand, various researches have proven that online travel reviews and posts help tourists make informed travel decisions. Additionally, these reviews and posts contain a lot of information that tourism managers and destination planners can utilize to improve their business. Online reviews and posts usually reflect the needs, wants, and expectations of the tourists, and understanding these factors is key in improving the business. However, very little research has been conducted concerning online reviews and their application in tourism development. Minkwitz[11] pointed out that studies about tourism and social media were mostly about using its influence on tourist behavior and the travel planning process, the use of UGCs in marketing tourism destinations, and creating image tourism establishments. Thus, this paper hopes to contribute to the literature on social media and tourism development planning.

This research aims to analyze traveler's comments and feedback about their Philippine travel experience to gauge the country's performance as a tourism destination and propose possible uses of UGC for tourism promotion and tourism development planning in the Philippines. To meet the objective of this study, three analysis methods will be conducted – VADER Sentiment Analysis, Word2Vec Analysis, and Cluster Analysis. The VADER Sentiment Analysis will be used to measure the overall sentiment of each review while the Word2Vec Analysis and Cluster Analysis will be used to analyze the review data and formulate proposals for the application of UGC in tourism promotion and development. Descriptive data is often analyzed using content analysis, but this research proposes the Word2Vec model as an alternative method in understanding and interpreting reviews for tourism development.

The succeeding section of this paper will provide an overview of the Philippine tourism industry. The next section describes the data collection process and provides a breakdown of the reviews and blog posts gathered. Section 4 will explain the analysis method employed in the study, followed by the presentation and discussion of the results. Section 6 contains the proposed application of the Word2Vec model output for tourism promotion and development and the findings' policy implications while Section 7 concludes this study.

2. PHILIPPINE TOURISM INDUSTRY

The Philippines is an archipelagic country located in the Southeast Asian region. With more than 7,000 islands, the country offers various activities for its tourists to enjoy. The National Tourism Development Plan (NTDP) 2016-2022[12] identified nine core tourism products (**Fig 1**) that the country can leverage on to increase its level of competitiveness and improve

Fig.1 Philippine core tourism products. Adapted from the Department of Tourism's National Tourism Development Plan 2016-2022

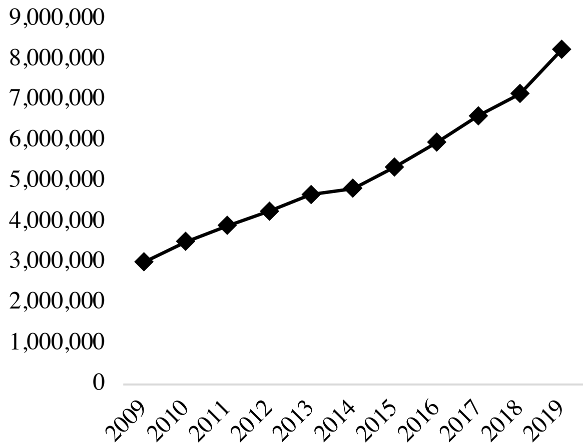

Fig. 2 Philippine international tourist arrivals (2009-2019)

the tourists' travel experience.

In the past years, the Philippine tourism industry has been experiencing continuous positive growth (**Fig 2**) with the majority of tourists coming from South Korea, China, the United States, Japan, and Taiwan[13]. The Visitor Sample Survey (VSS) [14] conducted by the Philippine Department of Tourism (PDOT) in 2019 showed that the top three Philippine destinations visited by tourists (excluding Manila) are Cebu, Boracay, and Palawan. These three destinations are considered the top island destinations in the Philippines and have been recognized by international tourism publications as some of the best islands in the world[15),16]. The same survey revealed that the top three tourist activities are shopping, sightseeing, and beach holiday.

Economically, the tourism industry has grown to be the fourth biggest industry in the Philippines. According to the Philippine Statistics Authority, tourism contributed 12.7% in the country's Gross Domestic Product (GDP) and employed 5.71 million people directly and indirectly in 2019[17], a 10.8% and 6.5% increase from the 2018 figures respectively, showing the industry's importance in the Philippine economy.

3. DATA GATHERING

The data used in this study were collected from three travel-related websites – TripAdvisor, Travellerspoint, and Off Exploring. TripAdvisor is the world's largest online travel community with more than 125 million travel reviews and opinions as of 2013[18] and more than 340 million monthly visitors and 78 million members as of 2015[19].Travellerspoint is another social networking website where people interested in travel can share their travel experience, get advice from fellow travelers, or simply form relationships with like-minded people[20]. Travellerspoint currently has more than 1 million members from 250 countries with around 500 new members per week[21]. Meanwhile, Off Exploring is a free travel blog service that enables its users to write about their travels and record their adventures in an online map. Off Exploring has 49,000 users as of September 2020[22].

A total of 1,717 Philippine travel reviews and blog posts from 39 countries written between 2000 and January 2020 were extracted from the abovementioned websites via Web Scraper[23], a free online web scraping browser extension. Information obtained from each review/post were username, year of posting, country of residence, review/post title, review/post body. The reviews and posts chosen for the analysis were those that talk about the Philippines as a tourist destination. Reviews and posts that focus on a specific tourism establishment or tourist attraction were excluded to ensure that the analysis will result in an overall impression on the Philippine travel experience.

The breakdown of the collected reviews/posts are shown in **Table 1** and **Fig 3**. More than half of the reviews/posts collected were written between 2011-2015 (54%). Further, the majority of the reviews/posts were written by travelers from the United Kingdom (31%), followed by Australia (12%), United States (10%), Philippines (7%), and Canada (6%). Travelers who did not share their country of residence also accounted for 22% of the dataset.

Table 1 Breakdown of review data by year of posting

Year Posted	Sample Size	Percentage Share
2000	4	0.23%
2002	1	0.06%
2005	3	0.17%
2006	34	1.98%
2007	60	3.49%
2008	81	4.72%
2009	100	5.82%
2010	99	5.77%
2011	213	12.41%
2012	137	7.98%
2013	232	13.51%
2014	186	10.83%
2015	156	9.09%
2016	95	5.53%
2017	143	8.33%
2018	90	5.24%
2019	69	4.02%
2020	14	0.82%
TOTAL	**1,717**	**100.00%**

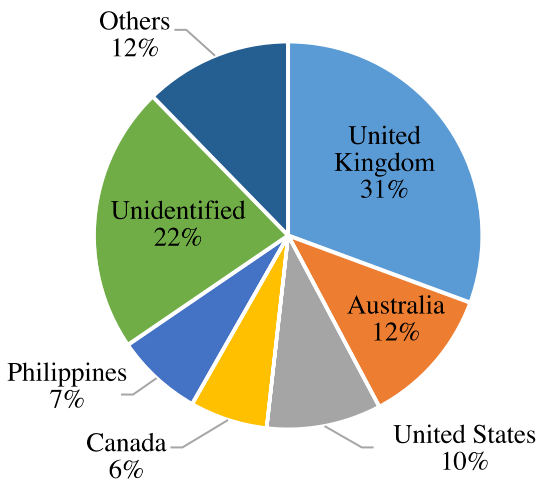

Fig. 3 Breakdown of review data by user's country of residence

4. DATA ANALYSIS & METHODOLOGY

This paper employed three analysis methods in analyzing the collected data – VADER Sentiment Analysis, Word2Vec Analysis, and Cluster Analysis.

(1) VADER Sentiment Analysis

VADER Sentiment Analysis is a sentiment analysis tool developed to address the challenge faced by traditional sentiment analysis in processing UGCs due to its inherent nature. Valence Aware Dictionary for sEntiment Reasoning (VADER) is a lexicon and simple rule-based sentiment analysis technique developed for social media texts. Unlike traditional sentiment analysis tools, VADER measures not only the polarity but also the intensity of the text based on the presence and/or absence of emoticons, punctuation marks, capitalizations, degree modifiers, contrastive conjunctions such as "but", and bi-gram and tri-grams[24]. The incorporation of these factors that are prevalent in the sentiment-oriented language of social media text thus makes VADER suitable for analyzing UGCs.

Python's VADER Sentiment module, a fully open-sourced software developed by Hutto and Gilbert[24] was used to conduct analysis. To preserve the integrity of the data, raw reviews and posts were used as input text in the module. The module then calculated the sentiment of each review/post and gave a compound sentiment score (CSS) between -1 (most negative) and +1 (most positive). For this study, reviews/posts with a CSS less than -0.5 were considered negative sentiment, CSS between -0.5 and +0.5 were treated as neutral, and reviews/posts with CSS higher than +0.5 were considered positive.

(2) Word2Vec Analysis

A team of Google researchers led by Tomas Mikolov developed the Word2Vec model in 2013[25]. It is a neural network that processes text data and generates a vector representation for each word in a multi-dimensional space. The vectors are then used to mathematically discover similarities among words.

Fig 4 shows how word vectors can be plotted in a two-dimensional space. It can be seen that the vector for the words “Queen” and “King” are close to each other in the same way “Princess” is to “Prince” and “Woman” is to “Man.” This means that the word “Queen” is most similar in meaning to “King” while “Princess” is most similar to “Prince” and “Woman” is most similar to “Man.” Likewise, based on the figure, it can be understood that “King” is more similar to “Prince” than it is to “Man.”

This example, however, only shows two dimensions. To be able to fully capture the meaning of the words, the vectors need to be calculated in multiple dimensions. Word2Vec calculates this using two learning models: Continuous Bag-of-Words (CBOW) and Skip- gram. The main difference between CBOW and Skip-gram is that CBOW predicts

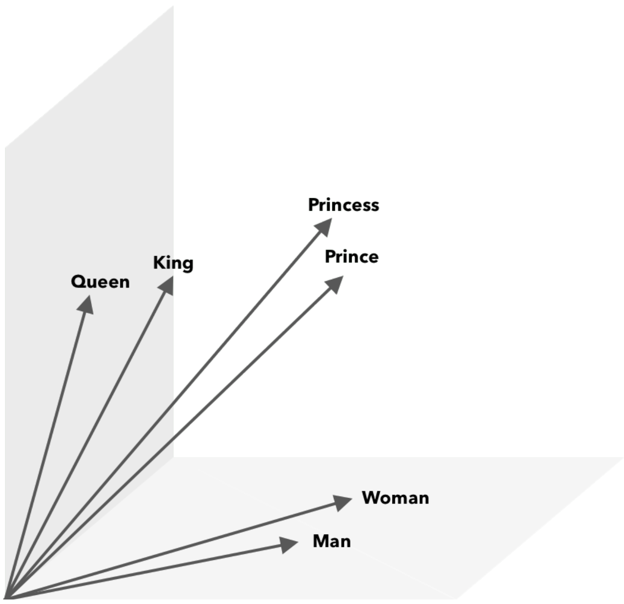

Fig. 4 Visual representation of word vectors in space. Adapted from “Topic Modeling With Word2Vec” by MarketMuse (2020)[26].

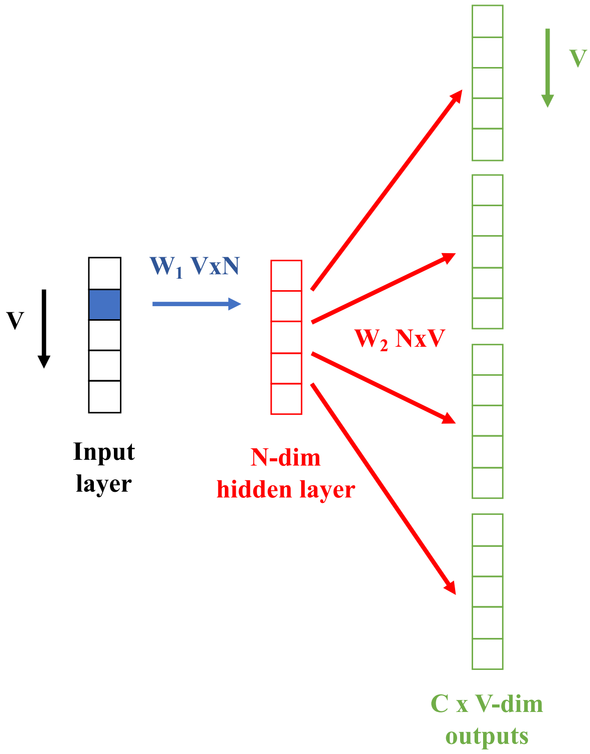

Fig. 5 Outline of word2vec’s Skip-gram module. Adapted from “Simple Tutorial on Word Embedding and Word2Vec” by Ali (2019)[27].

the target word based on the context of its surrounding words while Skip-gram uses the target word to predict the surrounding words. By feeding the text corpus into either CBOW or Skip-gram, the Word2Vec model builds a dictionary, learns each word’s vector representation across multiple dimensions, and finally generates word vectors that incorporate the semantic context of each word in the corpus[28]. This study used the Skip-gram module for the vector estimation and **Fig 5** is a visualization of how this module works. First, each word in the corpus is taken individually and assigned a vector for the input layer. Afterwards, by selecting a target word from the input layer and multiplying its vector by its weight matrix in N-dimensions in the hidden layer, a second weight matrix for the target word is calculated and then used to predict the target context words in the output layer. This process is done to all the words in the corpus until all words have their vector representation.

Aside from word vectors, Word2Vec also calculates for the cosine distance. The multi-dimensional characteristics of the word vectors gave each word multiple degrees of similarity and this degree of similarity among words is measured by cosine distance. Mikolov et.al showed the relationship between word vectors using the below example:

“What is the word that is similar to small in the same sense as biggest is similar to big?

Somewhat surprisingly, these questions can be *answered by performing simple algebraic operations with the vector representation of words. To find a word that is similar to small in the same sense as biggest is similar to big, we can simply compute vector X = vector ("biggest") − vector("big") + vector("small"). Then, we search in the vector space for the word closest to X measured by cosine distance…"*

In this study, Python's gensim library developed by Řehůřek, R., & Sojka, P.[29] was used to conduct the Word2Vec Analysis on the travel reviews and posts. Unlike the VADER Sentiment Analysis where raw reviews and posts were directly fed into the module and analyzed, Word2Vec Analysis via gensim has four steps (See **Fig 6**).

First, the reviews and posts were pre-processed and cleaned to ensure the accuracy of the model and the efficiency of data learning. The standard pre-processing procedure includes converting texts to lower case, removing unnecessary characters (e.g. numbers and punctuations), filtering stop words, tokenizing, and lemmatizing the vocabulary. However, the review dataset contains words and phrases that must be handled in a unique manner so three additional pre-processing procedures were conducted. The first one was correcting spelling errors. Afterwards, words that mean to the same thing but are called differently were integrated. An example of this would be the "Ninoy Aquino International Airport" which is also referred to as "NAIA" or "Manila airport." Considering that all three words refer to the same airport, the words "NAIA" and "Manila airport" were replaced with "Ninoy Aquino International Airport" for uniformity. Words and phrases like these were manually located and replaced. Lastly, compound proper nouns such as names of places (e.g. El Nido, Ilocos Norte, etc.) were joined with an underscore "_" so the program will treat it as one word. This process resulted in a vocabulary size of 22,017.

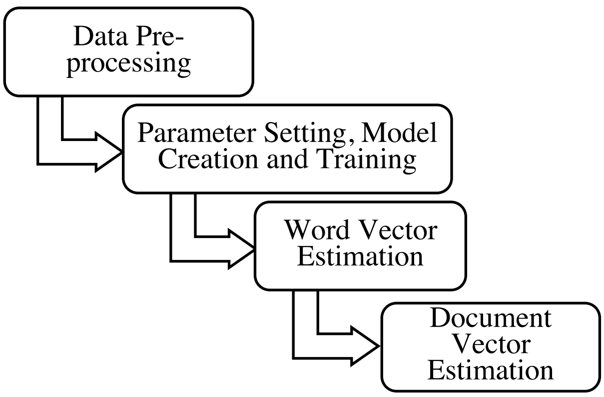

Fig. 6 Steps in conducting Word2Vec analysis in gensim

Table 2 Gensim parameters set for the Word2Vec analysis

Parameter	**Meaning**	**Study Parameter Value**
size	Dimensionality of word vectors	100
window	Maximum distance between the current and predicted word within a sentence	10
workers	Threads to train the model	4
min_count	Minimum frequency for words to be considered	1
sg	Training algorithm (0 = CBOW, 1 = skip-gram)	1
iter	Number of iterations (epochs) over the corpus	10

To facilitate data analysis, a Word2Vec model has to be created. For this study, the model parameters used for data processing was set up as shown in **Table 2**. Once the parameter was set, the model was then created, trained with the dataset, and word vectors for each vocabulary were estimated. Afterwards, the document vector of each review and post was estimated by calculating the average word vectors following the code developed by Spathis D.[30] and hosted on github.

(3) Cluster Analysis

To further understand the results of the VADER sentiment analysis and Word2Vec analysis, Cluster analysis was conducted. Cluster analysis is an unsupervised grouping technique used widely in various fields such as machine learning and image analysis to identify similarities within a dataset and group similar data together[31]. Cluster analysis is different from factor analysis in such a way that cluster analysis determines the grouping of the objects based on distance, while factor analysis bases its groupings on patterns of variation[32].

Cluster analysis has two main methods: hierarchical and partitional. In this research, hierarchical clustering, specifically the agglomerative approach, was conducted. Agglomerative hierarchical clustering treats each dataset as an individual cluster at the beginning before merging the two most similar clusters successively until there is only one cluster[33]. The cluster analysis was done with Python's Scipy[34] and Scikit-learn[35] libraries using the document vectors estimated with Word2Vec.

5. RESULTS AND INTERPRETATION

In this section, the results of all analyses are presented and discussed.

The first analysis conducted was the VADER Sentiment Analysis. Findings reveal that out of 1,717 reviews and posts, 1,516 are positive, 70 are neutral, and 131 are negative. This resulted in an average CSS of 0.7865, which is considered positive according to the threshold set at the beginning of the study. A detailed breakdown of the reviews shows that reviews from Canada have the biggest share of positive sentiments while reviews written by tourists from the United States have a lesser share (**Fig 7**). Nonetheless, more than 50% of the reviews in all countries are positive. **Fig 8** depicts the trend in the average CSS over time. It can be seen that average CSS for the years 2000, 2002, 2005, and even 2020 are very far from the average score. This could be due to the small number of reviews collected from those years,

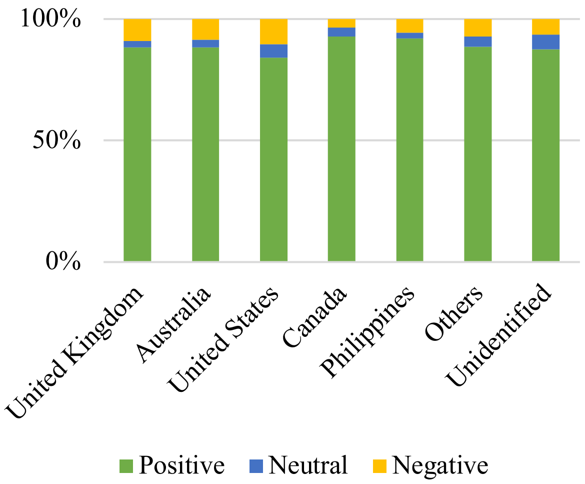

Fig. 7 Breakdown of review sentiments by user's country of origin

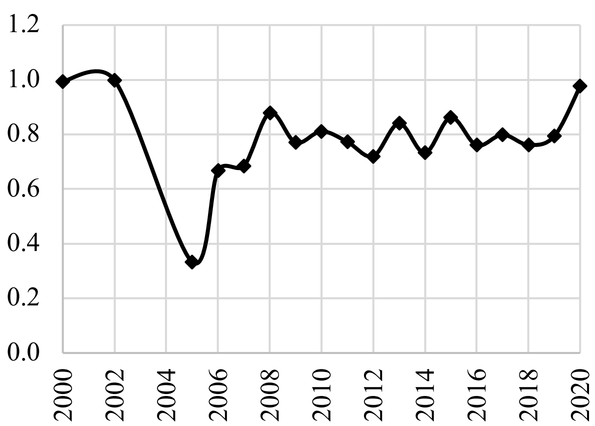

Fig. 8 Changes in average compound sentiment scores over time

affecting the average CSS. Thus, these CSSs can be considered outliers. If these scores are excluded, 2008 has the highest CSS with 0.8783 while 2006 had the lowest score of 0.6676.

Word2Vec analysis was conducted next and word and document vectors for the entire review dataset were estimated. The cosine distance of some of the relevant most frequent words ("manila", "beach", and "flight") are presented in **Table 3**. For the keyword "manila," most similar words that came up were names of country capitals and big cities (jakarta, hanoi, ho chi minh, and saigon), some areas in manila (baclaran and pasay), and words describing manila (metropolis, capital, and metro). The word "clark airport" is related to manila since it is the secondary airport nearest the capital and serves as an alternative to Manila's main airport, the Ninoy Aquino International Airport. The keyword "beach" returns popular beach destinations and beach resorts in the country,

Table 3 Top 10 most similar words based on Word2Vec estimate

Keyword	Most Similar Word	Cosine Distance
manila	jakarta	0.7079
	hanoi	0.7055
	baclaran	0.7033
	metropolis	0.7027
	ho chi minh	0.7011
	capital	0.6998
	saigon	0.6997
	metro	0.6988
	clark airport	0.6941
	pasay	0.6896
beach	diniwid beach	0.8010
	bounty beach	0.7772
	white beach	0.7716
	sand	0.7644
	baling hai beach resort	0.7625
	las cabanas	0.7510
	ilig iligan beach	0.7500
	overrate	0.7367
	sandugan beach	0.7334
	willys rock	0.7332
flight	cebu pacific airlines	0.8197
	airphil express	0.8049
	delay	0.7927
	airline	0.7886
	rebooked	0.7680
	ninoy aquino international airport	0.7634
	layover	0.7563
	pal express	0.7544
	airasia	0.7520
	clark airport	0.7444

while the words most similar to the keyword "flight" are airlines, airports, and flight-related occurrences such as "delay," "rebooked," and "layover."

Document vectors were then calculated by averaging the word vectors of each review and cluster analysis was conducted to divide the reviews into clusters for better understanding and interpretation. Due to the large number of reviews and posts, the dendrogram **(Fig 9)** that resulted from the cluster analysis was randomly cut in two varying distances to form a two-level nested cluster. Level 1 has a Euclidean distance of 3.5 and has seven "core" clusters while Level 2 has 31 "complementary" clusters with a distance of 1.75.

The reviews were then analyzed per cluster to identify each cluster's theme. **Table 4** shows some excerpts from reviews that fall under each core cluster and its corresponding complementary cluster. Reviews that were grouped under core cluster 1 include those that talk about the historical and cultural significance of a tourist attraction, and their experience with the Filipino hospitality. Core cluster 2 reviews were about marine life found in the Philippines and their experience when they tried some water activities. Travel tips such as how to get around and the different modes of transportation, and Philippine history were grouped into core cluster 3 while transportation-related issues and recommended accommodation were placed in core cluster 4. Core cluster 5 includes reviews that highlight the tourists' experience when they interacted with the locals during their travel and volunteer work while opinions and feedback about various aspects of their travel such as costs and accommodation were grouped into core cluster 6. Lastly, reviews about adventure activities such as trekking and spelunking in the Philippines were placed in core cluster 7.

The average CSS of each cluster was also computed. As can be seen in **Table 5**, all core clusters have positive sentiments, with reviews about Philippine beaches and water activities in Cluster 2 having the highest average CSS of 0.870793769 and impressions on Filipino people in Cluster 5 having the lowest average CSS of 0.579994872. By looking at this, one can assume that the Philippine tourism industry is doing well since all reviews written about the Philippine experience have positive sentiments. However, a closer look at the complementary clusters show otherwise. Out of 31 complementary clusters,three of them have neutral sentiments (Clusters 2, 6, 24) and one negative sentiment (Cluster 27). The neutral clusters contain reviews about airport/airline experience, transportation-related issues, and long bus rides. The only negative cluster contains one review about the Boracay food experience, which can be considered an outlier since there is only one review under that cluster.

The clustered reviews can be further broken down by the country of residence of its poster. Doing this shows that each market has its own characteristics.

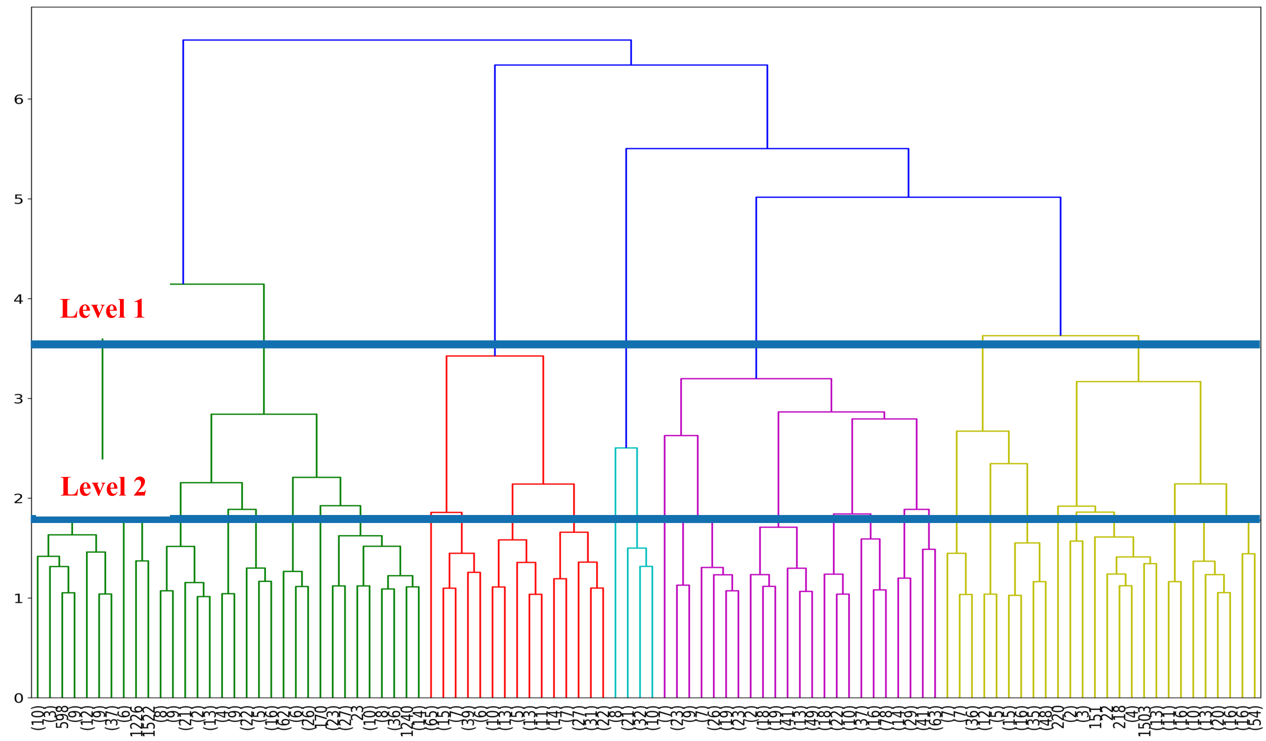

Fig. 9 Truncated dendrogram showing cluster identification of reviews/posts

Table 4 Excerpts from reviews in each cluster

CORE CLUSTER	COMPLEMENTARY CLUSTER	REVIEW EXCERPT
Core Cluster 1	Comp. Cluster 7	*"At certain parts of the road, Manila's two lane roads became a honking hive of vehicles. Jeepneys were the funky colorful vehicles used as one of the main transport for locals to get around the city, they where American vehicles left from the war"*
	Comp. Cluster 26	*"They could not speak any English but made us very welcome by offering us coffee and big smiles."*
Core Cluster 2	Comp. Cluster 17	*"The early morning brought us beautiful weather and the chance to swim with one of natures great animals the Whale Shark. We donned snorkelling gear and headed into an empty sea at the break of dawn to be greeted by these 3-5 metre behemoths with their gracious glide and gaping jaws, truly and amazing experience"*
	Comp. Cluster 22	*"Our guide paddled us out to the reef and we entered the water. The snorkelling was easy as we could just hold on to the outriggers of the boat"*
Core Cluster 3	Comp. Cluster 12	*"The most famous religious icon in the Philippines, the statue of Santo Nino, is said to have been given to Queen Juana by Magellan after her baptism and is kept in the oldest church in the country, the Basilica Del Santo Nino founded in 1565"*
	Comp. Cluster 23	*"How to visit Banaue and Batad was much more of a challenge than expected. First thought was to go by bus, but the only option appeared to be overnight…Have since read about suspension of service by some companies on this route (for licensing issues), so we are really glad we didn't opt for this…We then thought of renting a car and driving ourselves. Again, having been there do NOT do it… Eventually decided on a personalized tour with driver"*
Core Cluster 4	Comp. Cluster 6	*"From Dumaguete we caught another ferry to mainland Cebu before jumping on a bus for what would be one of the longest and most uncomfortable bus rides of my life. 5 hours cram packed on a tiny bus with no aircon and half a seat."*
	Comp. Cluster 9	*"I have stayed in Tamera Plaza Inn in Lacson Street. Rate ranges from P950.00 to P1500.00 for single accommodation and it is accessible in all vehicles plying the city. Rooms are clean and the ambiance will make you feel as if you're in a 5 star hotel."*
Core Cluster 5	Comp. Cluster 8	*"As soon as we landed Cebu we was greeted by our team leader, Samira, and a Filipino RV volunteer coordinator Karen"*
	Comp. Cluster 14	*"The children of Batad village could not have impressed me more."*
Core Cluster 6	Comp. Cluster 19	*"There were spots with many fishes, and were excellent for snorkelling. It costed only P2,500 - yes, that's it, just around US$55 for the whole boat, whole day!"*
	Comp. Cluster 25	*"Stayed at KOTA Beach resort. A nice family run property with air conditioned beach cottages"*
Core Cluster 7	Comp. Cluster 1	*"There are a lot of bats living in the cave, two different species of them. There are also cave swiftlets which we could hear coming and going"*
	Comp. Cluster 21	*"We left the small village of Banaue and started walking through the paths around the Ifugao terraces. We traversed around many hillsides giving great views of the terraces."*

Table 5 Theme and average compound sentiment score per cluster

CLUSTER	THEME	NO. OF REVIEWS	SENTIMENT SCORE
CORE CLUSTER 1	**Philippine Culture**	**429**	**0.781151981**
COMP. CLUSTER 7	Historical and cultural attractions	116	0.7482
COMP. CLUSTER 13	Food they ate in the Philippines	51	0.971876471
COMP. CLUSTER 16	Night life / Night activities	81	0.831874074
COMP. CLUSTER 3	Local interaction and criticisms	73	0.592126027
COMP. CLUSTER 26	Filipino hospitality	101	0.805735644
COMP. CLUSTER 30	Filipino cuisine	7	0.967285714
CORE CLUSTER 2	**Philippine Beaches and Water Activities**	**321**	**0.870793769**
COMP. CLUSTER 17	Marine life	99	0.923708081
COMP. CLUSTER 22	Water activities	113	0.888972566
COMP. CLUSTER 4	Quality of beaches	46	0.868206522
COMP. CLUSTER 18	Diving	63	0.756925397
CORE CLUSTER 3	**Overview of the Philippine Tourism**	**215**	**0.838106977**
COMP. CLUSTER 11	Good experience with a tour guide	9	0.816255556
COMP. CLUSTER 12	Philippine history	53	0.626669811
COMP. CLUSTER 23	Getting around / Modes of transportation	37	0.938462162
COMP. CLUSTER 5	Good experience in the Philippines	17	0.850147059
COMP. CLUSTER 15	Island destinations	75	0.954674667
COMP. CLUSTER 20	Bohol attractions	23	0.795682609
COMP. CLUSTER 28	Enjoyed Boracay	1	0.5563
CORE CLUSTER 4	**Logistics and bookings**	**241**	**0.582649378**
COMP. CLUSTER 24	Long bus ride	2	-0.17
COMP. CLUSTER 9	Accommodation used and recommendations	102	0.75745098
COMP. CLUSTER 10	Trip-planning recommendations	32	0.59661875
COMP. CLUSTER 2	Airport/Airline experience	26	0.198434615
COMP. CLUSTER 6	Transportation-related issues	79	0.496802532
CORE CLUSTER 5	**Impressions on Filipino people**	**78**	**0.579994872**
COMP. CLUSTER 14	Interaction with the locals	54	0.594796296
COMP. CLUSTER 8	Volunteer work in the Philippines	24	0.546691667
CORE CLUSTER 6	**Comments and Feedbacks**	**261**	**0.862944828**
COMP. CLUSTER 19	Comments on costs and expenses	74	0.903095946
COMP. CLUSTER 25	Feedback on accommodations	98	0.950053061
COMP. CLUSTER 27	Boracay food experience	1	-0.4767
COMP. CLUSTER 29	Comments about traveling within the Philippines	74	0.784860811
COMP. CLUSTER 31	Mix hotel reviews	14	0.549378571
CORE CLUSTER 7	**Adventure activities**	**172**	**0.840792442**
COMP. CLUSTER 21	Trekking and Climbing	66	0.899916667
COMP. CLUSTER 1	Spelunking and other nature activities	106	0.803979245

For example, **Fig 10** shows that tourists from the UK mostly write about their visit to historical and cultural attractions, and the nightlife in the area they visited. On the other hand, tourists from Australia mostly write about diving and the accommodation they stayed in while American tourists talk about their accommodation and their interaction with the locals. Reviews written by those who live in Canada are mostly about water and other nature-based activities, and local tourists can be considered the most cost-conscious travelers as the majority of their comments are about the expenses, followed by historical narratives.

6. APPLICATION OF WORD2VEC OUTPUT IN TOURISM PROMOTION AND DEVELOPMENT AND ITS POLICY IMPLICATIONS

This section provides practical applications of the output from the word2vec analysis for tourism promotion and development and the output's tourism policy implications.

First, **Table 5** can be used as a reference to identify areas that need improvement to provide tourists with better tourism experience. The word2vec and VADER sentiment analysis revealed that tourists have neutral sentiments about airport and airline experience, other transportation-related issues, and long bus rides. A deeper look into these clusters revealed that tourists mostly complain about road congestion and traffic, over-pricing incidents, schedule delays and cancellations, and bad customer service. Therefore, these concerns must be addressed to improve the tourists' experience. Since all three neutral clusters talk about transportation, one recommendation would be to strengthen the Department of Tourism's coordination with the Department of Transportation and come up with a policy that will ensure that better transportation options and improved transportation facilities are offered to tourists.

Additionally, **Fig 10** reveals each market's characteristics and what they value in a tourist destination. This information can be utilized in positioning the Philippines better in the market. For example, since reviews from the UK are about historical and cultural Philippine attractions, the country can be promoted as a cultural destination in the UK. Likewise, the Philippines can be promoted as a diving destination for Australians and a sun and beach getaway for the Canadians. Visitors from the US, on the other hand,

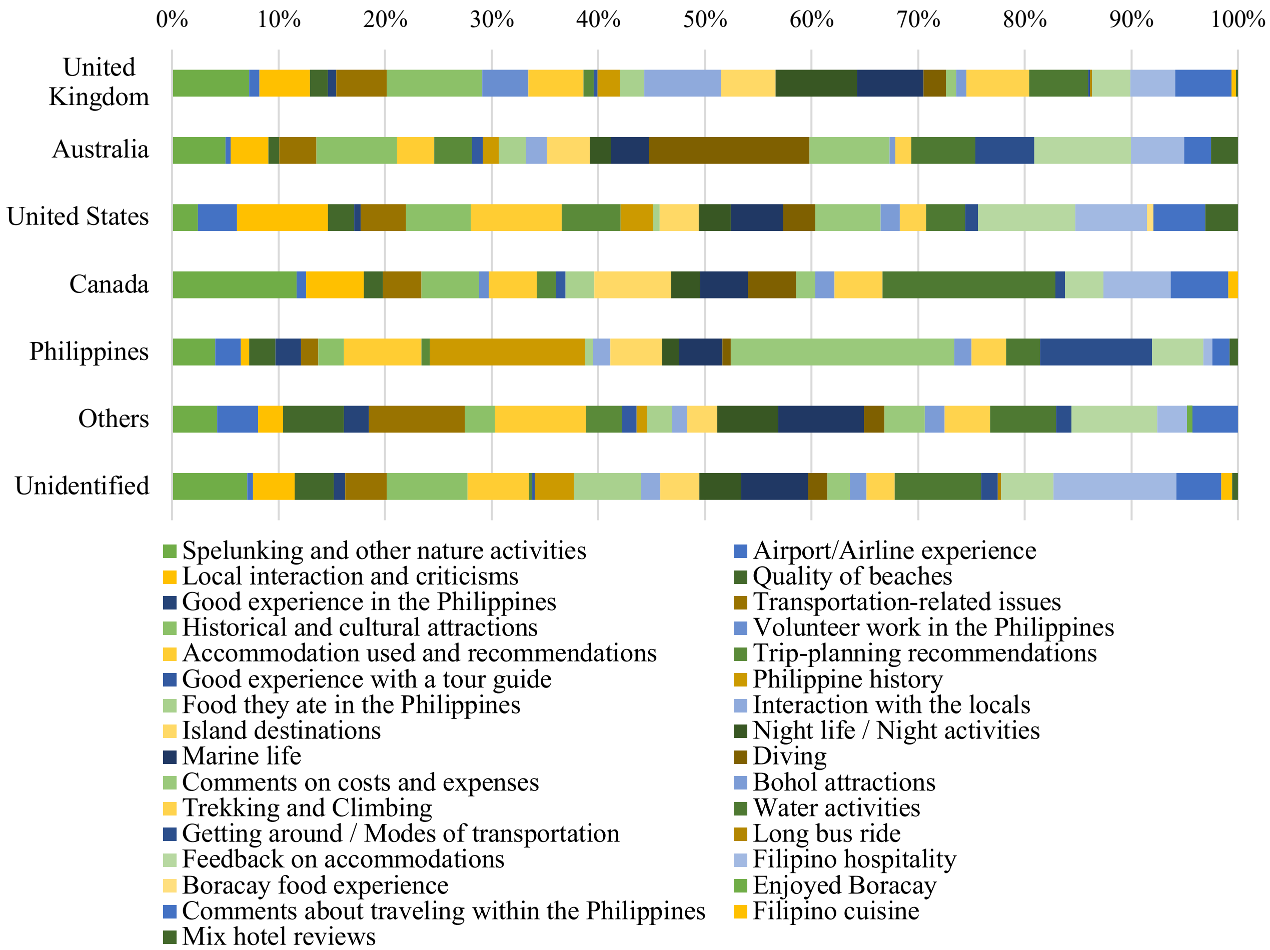

Fig. 10 Cluster breakdown of reviews by country of residence of poster

usually talk about local interaction. This could be due to the fact that there are about 4 million Filipino migrants and Americans with Filipino ancestry in the US[36] and the American tourists who come to the Philippines do so to visit their Filipino relatives and friends. The Department of Tourism can capitalize on this knowledge to identify which products to prioritize for development, promotion, and preservation. For instance, the cultural and historical resources of the country should be properly maintained and preserved to make sure that it meets the standards of the British tourists. In the same way, development of the country's natural resources needs to be controlled so that Australian and Canadian tourists who visit the Philippines for diving and other nature-based activities will have a pleasant experience. Meanwhile, Filipinos like to write about Philippine history so domestic tourism can be promoted by highlighting the historical significance of the destination and offering affordable packages to locals.

Lastly, the word vectors can be used to find out potential alternative destinations and activities by calculating the distance between vectors. **Table 6** shows some of the possible word vector combinations that can be useful in tourism development. In example 1, the vectors were used to look for a possible alternative destination for people who goes to Boracay (Word A) for its White Beach (Word B), and the closest word to the vectors of "Boracay" and "White Beach" is Mindoro with a distance of 0.44. **Fig 11** illustrates the geographical orientation of Boracay and Mindoro and an image of Boracay's White Beach and a beach in Mindoro that can be promoted as an alternative. The vectors can also be used to discover destinations that offer similar experiences. Example 2 shows that Bontoc can be offered in place of Bohol for tourists who wish to experience Philippine culture while in example 3, the vectors of two types of tourist activities were used to come up with a recommended destination where both activities can be done. A possible tour combination can also be suggested in place of one single destination. In example 4 and **Fig 12**, Dumaguete and Panglao Island can be promoted as a tour package to tourists who wish to visit Mindoro to dive. Interestingly, this aspect of word2vec is what makes it different from content analysis and other qualitative data analysis methods. The conversion of texts to vectors allows for this simple addition (and subtraction) to be done on the data, providing a considerable number of word combinations to work with, depending on the user's needs, thus proving the word2vec model's advantage over qualitative analysis methods.

	Word A	**Word B**	**Closest Word (Distance)**
Ex. 1	Boracay	White Beach	Mindoro (0.44)
Ex. 2	Bohol	Culture	Bontoc (0.82)
Ex. 3	Trek	Dive	Coron (1.32)
Ex. 4	Dumaguete	Panglao Island	Mindoro (1.84)

Table 6 Sample word vector combinations and closest word based on distance

Based on the findings presented in the previous chapter and the applications of the word2vec discussed in this section, it was established that travel reviews and posts contain abundant information that can be used for tourism planning and development, and that the word2vec model is a powerful tool that can be used to analyze descriptive data. Thus, as a policy recommendation, the Department of Tourism needs to invest and capitalize on the power of social media and user-generated contents by making sure that all tourism-related establishments have an online presence. Doing so will not only widen their audience and boost their sales[37], the information that can be gathered from the reviews has proven beneficial for business improvement and local tourism development as well.

7. CONCLUSION

In this paper, online travel reviews and blog posts about the Philippine travel experience were collected and analyzed to gauge the Philippines' performance as a tourist destination and to use the reviews and posts to further develop tourism in the country. VADER Sentiment Analysis, Word2Vec Analysis, and Cluster Analysis were conducted on the gathered data. VADER Sentiment Analysis was employed instead of traditional sentiment analysis due to VADER's attunement to the language used in social media, thus making it the more appropriate analysis method for this dataset. This analysis revealed that overall, tourists express a positive sentiment towards their experience in the Philippines. Word2Vec Analysis and Cluster Analysis were then conducted to further analyze the review data for its application in tourism development and promotion. The document vectors estimated using word2vec resulted in 7 core clusters and 31 complementary clusters, with all core clusters having positive sentiments and 4 out of 31 complementary clusters having neutral and negative sentiments. The clusters were also broken down per country of residence of the posters and the results showed the characteristics of each market.

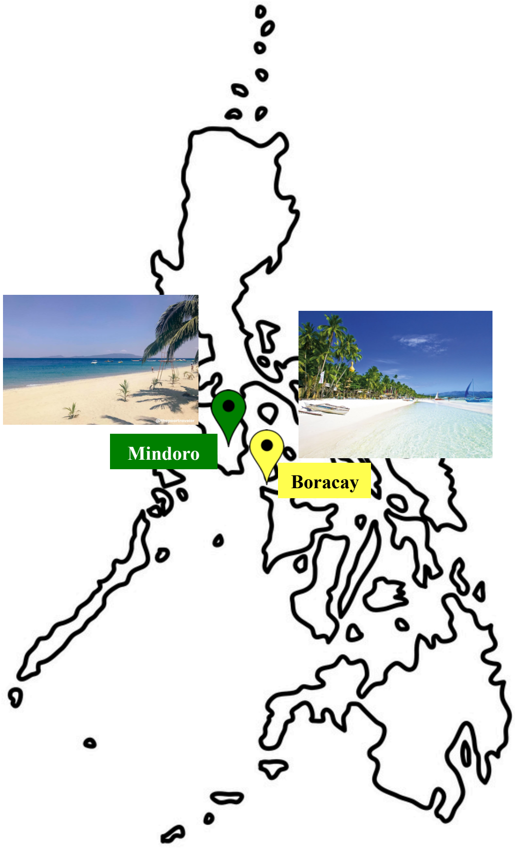

Fig. 11 Geographical orientation of Boracay and Mindoro and images of Boracay and Mindoro beaches

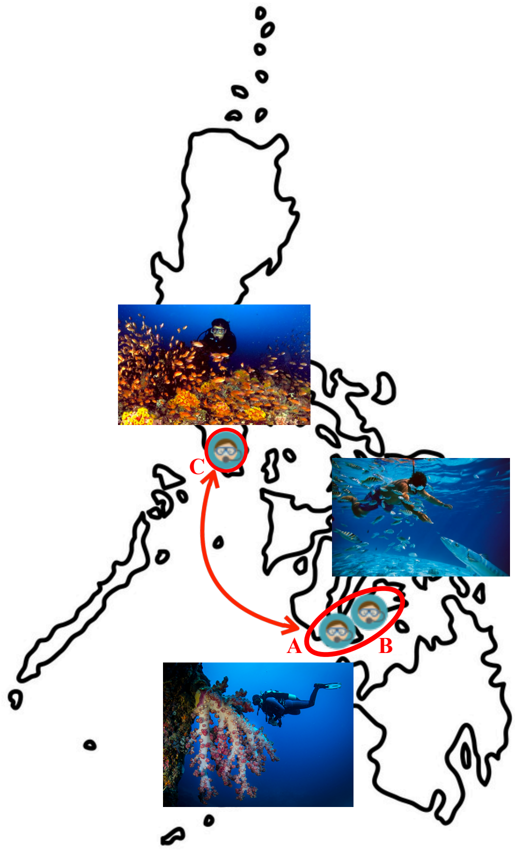

Fig. 12 Diving in Mindoro can be substituted by diving in Panglao and Dumaguete

The possible applications of the findings in tourism development and promotion were also discussed. First, the insights gathered from the reviews can be used to identify areas of improvement to better the tourism experience in the Philippines. It can also be used as a reference when crafting national tourism development plans and strategizing tourism promotional plans by having an understanding of what each market wants and what they look for in a tourism destination. Lastly, simple addition and subtraction can be applied to the word vectors estimated from the word2vec analysis to identify alternative destinations and tour combinations.

As for policy implications, this study has proven that abundant information applicable to tourism development can be gathered from travel reviews and travel blogs that are easily available online. Therefore, it is only normal to encourage all tourism businesses to establish their own online presence. To do this, policies supporting the provision of an internet connection should be put in place, especially those in rural areas since they have the tendency to suffer from internet divide[38) compared to urban areas. Manpower training should also be provided to ensure that tourism businesses know how to fully utilize the information available online. Furthermore, the analysis revealed that reviews and posts with low compound sentiment scores were about transportation. Thus, closer coordination between the Philippines' transportation and tourism sector is needed for transport policy and tourism policy to be more collaborative to enhance the transportation facilities in the country.

REFERENCES

1) Delener, N.: Current trends in the global tourism industry: evidence from the United States, *RAP – Rio de Janeiro*, Vol. 44, pp. 1125-1137, 2010.
2) Guerreiro, C., Cambria, E., & Nguyen, H.T.: Understanding the Role of Social Media in Backpacker Tourism, *2019 International Conference on Data Mining Workshops (ICDMW)*, pp. 530-537, 2019.
3) Ráthonyi, G., & Varallyai, L.: Web 2.0 and Tourism, *Proceedings of the International Conference on Information and Communication Technologies for Sustainable Agri-production and Environment (HAICTA 2011)*, pp. 27-39, 2011.
4) World Tourism Organization: Online Guest Reviews and Hotel Classification Systems – An Integrated Approach, UNWTO, Madrid, 2014.
5) Browning, V., So, K.K., & Sparks, B.: The Influence of Online Reviews on Consumers' Attributions of Service Quality and Control for Service Standards in Hotels, *Journal of Travel & Tourism Marketing*, Vol. 30, pp. 23-40, 2013.
6) Chong, A., Khong, K.W., Ma, T., McCabe, S., & Wang, Y.: Analyzing key influences of tourists' acceptance of online reviews in travel decisions, *Internet Res.*, Vol. 28, pp. 564-586, 2018.
7) O'Connor, P.: User-Generated Content and Travel: A Case Study on Tripadvisor.Com, In *Information and Communication Technologies in Tourism 2008*. Springer, Vienna, 2008. https://doi.org/10.1007/978-3-211-77280-5_5

8) Li, X., & Hitt, L.: Self Selection and Information Role of Online Product Reviews, *Information Systems Research*, Vol. 19, pp. 456-474, 2008.
9) Jayathilaka, R., Dharmasena, T., Rezahi, N., & Haththotuwegama, S.: The impact of online reviews on inbound travellers' decision making, *Quality & Quantity*, Vol. 54, pp. 1005-1021, 2020.
10) Alaei, A., Becken, S., & Stantic, B.: Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, Vol. 58, pp. 1-17, 2017.
11) Minkwitz, A.: TripAdvisor as a Source of Data in the Planning Process of Tourism Development on a Local Scale, *Turyzm*, Vol. 28, pp. 49-55, 2018.
12) Department of Tourism: National Tourism Development Plan 2016-2022.
13) Department of Tourism: Tourism Demand Statistics, Accessed 11 September 2020 from http://www.tourism.gov.ph/.
14) Department of Tourism: Visitor Sample Survey 2019.
15) Marino, V.: The Best Islands in the World: 2019 Reader's Choice Awards, *Conde Nast Traveler*, 2019, Accessed 11 September 2020 from https://www.cntraveler.com/.
16) Clemence, S.: The Top 25 Islands in the World, *Travel + Leisure*, 2020, Accessed 11 September 2020 from https://www.travelandleisure.com/.
17) Philippine Statistics Authority: Philippine Tourism Satellite Accounts, 2020, Accessed 07 September 2020 from https://psa.gov.ph/.
18) Yoo, K., Sigala, M., & Gretzel, U.: Exploring TripAdvisor, In *Open tourism*, pp. 239-255, Springer, Berlin, Heidelberg, 2016, https://doi.org/10.1007/978-3-642-54089-9_17.
19) Bhatnagar, U.: Analysing the Effect of TripAdvisor on Reservation Intention of Hotels for Travellers from New Delhi, India, *Dusit Thani College Journal*, Vol. 12, pp. 30-49, 2018.
20) Cai, J., Qiu, L., & Ren, F.:. A Service Integration Model of Value Creation: A Study of Commercial Online Communities, *AMCIS 2008 Proceedings*, 2008, https://aisel.aisnet.org/amcis 2008/327.
21) Travellerspoint.com: Travellerspoint Stats, Accessed 07 September 2020 from https://www.travellerspoint.com/.
22) Offexploring.com
23) Webscraper.io
24) Hutto, C.J. & Gilbert, E.E.: VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI, 2014.
25) Mikolov, T., Chen, K., Corrado, G.S., & Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR*, 2013, abs/1301.3781.
26) Jeske, S.: Topic Modeling With Word2Vec, MarketMuse, 2020, Accessed 26 September 2020 from https://marketmuse.com/
27) Ali, Z.: Simple Tutorial on Word Embedding and Word2Vec, Medium, 2019, Accessed 26 September 2020 from https://medium.com/
28) Ma, L., & Zhang, Y.: Using Word2Vec to process big text data, *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2895-2897, Santa Clara, CA, 2015.
29) Řehůřek, R., & Sojka, P.: Software framework for topic modelling with large corpora, *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45-50, 2010.
30) Spathis, D.: Average words to represent documents with Word2Vec, Github, San Francisco, CA, 2016, Accessed 16 June 2020 from https://github.com/sdimi/average-Word2Vec.
31) Omran, M.G., Engelbrecht, A., & Salman, A.: An overview of clustering methods, *Intelligent Data Analysis*, Vol. 11, pp. 583-605, 2007.
32) Setyaningsih, S.: Using Cluster Analysis Study to Examine the Successful Performance Entrepreneur in Indonesia, *Procedia. Economics and finance*, Vol. 4, pp. 286-298, 2012.
33) Ackermann, M., Blömer, J., Kuntze, D., & Sohler, C.: Analysis of Agglomerative Clustering, *STACS*, pp. 308-319, 2011.
34) Jones, E., Oliphant, T. & Peterson, P.: SciPy: Open Source Scientific Tools for Python, 2001.
35) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, D., Brucher, M., Perrot, M., & Duchesnay, E.: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
36) INQUIRER.NET U.S. Bureau: Filipino population in U.S. now nearly 4.1 million — new Census data, *Inquirer.net*, 15 November 2019, Accessed 16 September 2020 from https://usa.inquirer.net/
37) Kumar, S.: Internet Marketing in Tourism, *Indian Journal of Applied Research*, Vol. 4, pp. 286-289, 2014.
38) Nedescu, D.: In the Philippines, the urban-rural 4G Availability divide varies by region, *OPENSIGNAL*, 29 October 2019, Accessed 17 September 2020 from https://www.opensignal.com/

(Received October 1, 2020)