
IFC-Bank of Italy Workshop on “Data Science in Central Banking: Applications and tools”

14-17 February 2022

Sentiment analysis of tourist reviews from online travel forum for improving Indonesia tourism sector¹

Muhammad Abdul Jabbar, Arinda Dwi Okfania,
Anggraini Widjanarti and Alvin Andhika Zulen,
Bank Indonesia

¹ This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Sentiment Analysis of Tourist Reviews from Online Travel Forum for Improving Indonesia Tourism Sector

Muhammad Abdul Jabbar¹, Anggraini Widjanarti², Alvin Andhika Zulen³, Arinda Dwi Okfantia

Abstract

Tourism is one of strategic sector in Indonesia to increase Gross Domestic Product (GDP). In the long run, tourism sector can support exchange rate management and reduce current account deficit. Therefore, it is pivotal for optimizing tourism sector by improving aspect that is important for tourist. Text sentiment analysis, has become a flourishing frontier in the text mining community. This project studies tourist reviews of tourist destinations on the online travel forum, TripAdvisor, to produce 5 (five) tourism indicators, namely: the composition of foreign and domestic tourists, the composition of the country of origin of tourists, the movement of visits from month to month, the rating of tourist attractions, and positive and negative sentiments of tourists towards tourist destinations in the 3A aspect (attractions, amenities, and accessibilities). These five indicators can be used to help build a promotional strategy that is targeted and effective. The results of this positive review can be used to strengthen aspects of tourist destinations that can be promoted. Meanwhile, the negative review results are used as recommendations for future improvements in tourist attractions.

Keywords: text mining, sentiment analysis, tourism.

JEL classification: L83, L86

¹ Statistics Department – Bank Indonesia; E-mail: Muhammad_abdul@bi.go.id

² Statistics Department – Bank Indonesia; E-mail: Anggraini_widjanarti@bi.go.id

³ Statistics Department – Bank Indonesia; E-mail: Alvin_az@bi.go.id

Contents

1. Background.....	3
2. Literature Review	4
2.1 Tourism Quality Aspects	4
2.2 TripAdvisor Review Reliability	5
2.3 Text Mining on Tourism	6
3. Methodology.....	6
3.1 Data	6
3.2 Framework	8
Data Pre-processing	8
Sentiment Analysis	9
4. Result & Analysis	10
4.1 Tourists Country of Origins Composition	10
4.2 Tourism Destinations Monthly Visits	12
4.3 Tourism Destinations Rating	13
4.4 Sentiment Analysis Results.....	14
5. Conclusion & Future Works	19
5.1 Conclusion	19
5.2 Future Works	20
References.....	21

1. Background

Since 2011, Indonesian Government has launched tourism promotion program that has been using "Wonderful Indonesia" as main tourism branding tagline. As one of the strategic economic sector with big multiplier effect, tourism has been one of the sector that highly supported from the government. It has shown with many initiatives from the government to accelerate the development. In 2019 the government has set a high target of 20 Million foreign tourists and foreign exchange earnings of 17.6 Million USD with tourism shares of GDP around 5%. The improvement and development of major tourism destinations and events has been done steadily to realize Indonesia as world top tourism destinations.

Continuous and rigorous tourism sector development is necessary to support the growth and resilience of Indonesia external economy. The tourism sector supports Indonesia foreign exchange earnings and current account balance. To achieve Bank Indonesia vision to create a tangible national economic contribution, Bank Indonesia is closely coordinating with the government to support policy strategy formulation of tourism sector development in Indonesia.

The fact is, in Indonesia, Tourism sector data is relatively hard to find. Other than number of tourists that obtained from immigration, most of the data is acquired via survey. One of useful data that can be used in tourism sector development is the fulfilment of 3A aspects (attraction, amenities, and accessibilities) in Indonesia tourism destinations. This information can be used by regional government to provide recommendation and feedback regarding the development of each of tourism destinations on their region. However, this information is difficult to obtain in a bigger scale and usually done via survey that may need a lot of time and high cost in its collection.

Along with the growth of internet and smartphone technology, there are more online travel platform such as TripAdvisor, Booking.com, Agoda, & Traveloka. With the microblogging trends where users voluntarily and happily provide reviews of their travels, this drive the growth of user generated data in these platforms. These online travel platforms gathered large amount of user generated data that can be useful for many kinds of analysis including sentiment analysis. The Big Data that available on the platforms has big potential not only for the platforms, but also for government, authorities and tourism destination providers.

A study by Pan in 2007 shows that online reviews of products and services including tourism are deemed more trustworthy and credible. Recommendation based on previous user experience is one of the most influential information for the reader travelling decision. With the advancement of computer technology, we can extract information from large amount of text using Big Data Analytics and text mining methodology. These technology can be used to develop new indicator in the tourism sector that is credible and timely.

2. Literature Review

2.1 Tourism Quality Aspects

Tourism development involves components regarding tourism planning approaches, tourism attractions, accommodation, tourism facility and other tourism services such as transportation, infrastructure, and tourism management institutions (Inskeep, 1991). Tourism attractions can be categorized into 3 categories:

1. *Natural attraction* is attraction based on natural environment formation. Examples of natural attractions are climate, scenery, flora and fauna and also other form of nature uniqueness.
2. *Cultural attraction* is attraction based on people cultural activities. Examples of cultural attractions are archeological attractions, religious attractions, and traditional lifestyle.
3. *Special types of attraction* are man-made attraction such as theme park, circus, shopping mall, etc.

From (Yoeti, 1997) studies we understand that several aspects that correlate with the success of a tourism destination are 3As, the destination attractiveness (attraction), how the ease of access to the destination (accessibility) and the facility destination (amenities). Similarly, (Middleton, 2001) and (Mason, 2003) defines tourism product is the combination of 3 main components of attractions, the available facilities, and the destination accessibilities.

First, attraction is the elements in a tourism destination that broadly determine consumer or tourist's choice of what kind of attraction that they want to see or visit. The categories of attractions are natural tourism (involving nature, beach, climate, and other geographical and natural resources), man-made attractions (involving building and infrastructure such as historical and modern architecture, monument, garden, and thematic area), cultural attractions (involving historical and folklore, art and religion, musical theatre, museum, and other special events), and social attractions (involving the life local habitants within an area, language, and other social activities).

Second, amenities or the facility within a tourism destination is the elements that enables the tourists to participate and stay to enjoy the attraction of a tourism attraction. These elements are accommodation (hotel, apartment, villa, caravan, hostel, and guest house), restaurant, transportation (taxi, bus), other facilities (toilet and religious spaces), retail outlet (shops, travel agent, souvenir), and other services.

Third, accessibilities is the ease of access to the tourism destination that involves the cost, smoothness and comfort of travel through the infrastructure, road, airport, railway, seaport, operational factors such as operational route, transportation service frequency and cost, and government regulation regarding transportation safety.

Indonesia ministry of tourism also states that tourism development involves 3 aspects:

- 1) **Attractions**, site attractions (historical sites, places with comfortable or delightful climate, and places with beautiful sceneries), and event

attractions (eventful occasion such as congress, exhibition, or other kind of events);

- 2) **Amenities**, the availability and quality of the facilities, such as: Inn, restaurant, local transportation, and communication services;
- 3) **Accessibility**, the duration, cost, comfort, security, and availability of transportation to the location.

2.2 TripAdvisor Review Reliability

As the participatory nature of the Internet permits easy contribution of user-generated content without any editorial control, sites such as TripAdvisor allow users to share post-trip travel experiences have acquired immense popularity (Lo, 2011). TripAdvisor users can post review without any verification of whether the information that they submitted is factual or not and mostly based on their own experienced that can be very subjective, therefore raise some concerns of its reviews veracity.

Chua's study uses Social Network Analysis (SNA) to form clusters of highly-interlocked hotels that have been evaluated by a common pool of reviewers. The results of SNA indicated 58 isolated hotels that were not evaluated by any reviewer who had also commented on some other hotel in Singapore. These included 39 two-star hotels, 17 three-star hotels and two four-star hotels. The remaining 191 hotels (249 - 58) were interlocked cumulatively accounting for 4,192 interlocks with one another. The highly-interlocked hotels with five or more interlocks were used for further analysis. There were 46 such hotels accounting for 310 interlocks with one another. The hotels included three three-star hotels, 23 four-star hotels and 20 five-star hotels. Reviews posted by the top 100 reviewers in terms of volume of entries for the selected 46 interlocked hotels were used for further investigation. The top 100 reviewers contributed 424 reviews with an average of 8.83 reviews per hotel.

On establishing review baselines, the inter-reviewer and intra-reviewer reliability of contributions was studied using reviewer-centric and review-centric approaches respectively. Results suggest that reviews in TripAdvisor could be largely reliable (Chua, 2013).

SNA Result Illustration (Chua, 2013)

Figure 1

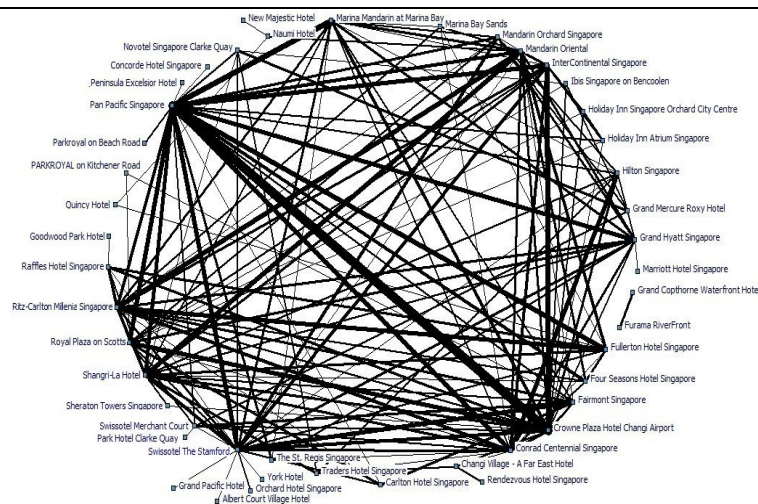


Fig. 1. Interlocking patterns of the 46 hotels with more than five connections.

2.3 Text Mining on Tourism

Text mining as defined by Hearst (2003) is the process of deriving high-quality information from text. It involves the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. In text mining, the goal is to discover heretofore unknown information, something that no one yet knows and so could not have yet written down.

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity (Medhat, 2014). Sentiment Analysis usually done by deriving information from large amount of data to studies positive and negative perceptions regarding an object. Sentiment Analysis usually done by creating classification models that can accurately predict whether a sentences expresses positive or negative sentiment toward the object that want to be analyzed.

Text mining is widely used in a lot of field, including tourism. Park utilizes text mining and twitter data to help formulate strategies of yacht cruise promotion (S.B Park, 2016). Text mining also utilized in hospitality industry to classify customer satisfaction from customer reviews (Berezina, 2016). A study by Pan states that online reviews of products and services including tourism are more accurate and timely (Pan, 2007). Recommendation based on experience of other users is one of the most influential source of information for a traveler's decision.

Thaha did a study that extract information from tourist reviews of Tangkuban Perahu and Kawah Putih in Indonesia (Thaha, 2020). The study extract the positive and negative reviews on the two tourism destinations from the visitors of the destinations. The result shows the rating of the tourism destinations and the positive sentiments for the rating, also negative reviews with words such as "expensive". The study uses Ekman emotion classification which is a classification of cross-cultural basic emotions proposed by Ekman et.al in 1992. These emotions are anger, disgust, fear, happiness, sadness, and surprise. The study shows that the review is dominated by "joy" emotion which indicates that most tourist enjoy the destination. From the negative sentiments, they extract "sadness" and "anger" emotion, and the details of the reviews can be used for recommendation by the tourism destination management.

3. Methodology

3.1 Data

Indonesia tourism destinations has similar attraction values to those in Thailand and Vietnam. The three peer countries tourism destinations strengths are its natural attractions and the sense of South East Asian spirituality and culture. Indonesia tourism destinations are competing with Thailand and Vietnam destinations since both countries have the same category of attractions and tourists market shares, especially for foreign tourists from Japan, China and Australia. Therefore, to analyze Indonesia tourism destinations competitive value, we compare it with similar tourism destinations in Thailand and Vietnam.

This study uses tourist reviews data that are extracted from TripAdvisor.com using its API. TripAdvisor API is a powerful API that enable its user to extract large amount of user reviews with all of its information in its database with specifiable parameters of tourism destinations or locations. This study used TripAdvisor Content API with specified parameters of tourism destinations in the regions we want to observe. The reviews are extracted from "things to do" pages for tourist destinations Indonesia, Thailand and Vietnam. The pages are crawled to extract its reviews automatically. This methodology extracted 604,395 reviews from 1330 tourist places within 26 major tourism destinations in Indonesia, Thailand and Vietnam from the year 2004 to 2021.

The information that extracted from the reviews are:

1. Tourists Destination Name
2. Category (Nature parks, museum, etc.)
3. Address
4. Rating information (Aggregates and number of reviews per rating scales(excellent, very good, average, poor, terrible)
5. Review information (Reviewer name, country of origin (not mandatory), review date, review content, visit date, rating, language)

The list of tourists destinations are as follows:

List of Tourism Destination

Table 1

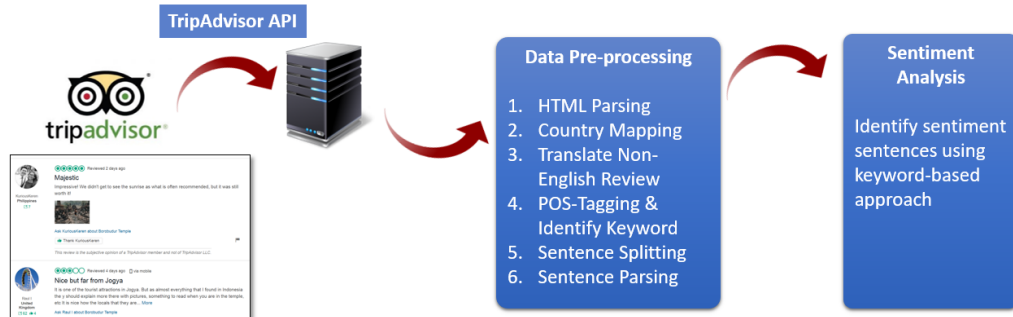
Indonesia		Thailand	Vietnam
Borobudur	Wakatobi	Chiang Mai	Ho Chi Minh City
Bromo	Morotai	Bangkok	Halong Bay
Banyuwangi	Danau Toba	Phuket	Nha Trang
Kepulauan Seribu	Tanjung Kelayang	Ayutthaya	Hoi An
Kota Tua	Bali		
Yogyakarta	Solo		
Tanjung Lesung	Semarang		
Mandalika	Batam		
Labuan Bajo	Likupang		

3.2 Framework

The text mining framework of the reviews extraction is as follows:

Sentiment Analysis Framework

Figure 2



Data Pre-processing

Pre-processing is done to convert unstructured html data into structured data and enrich the information of each reviews prior to the sentiment analysis of reviews. The pre-processing steps in this paper are as follows:

1. HTML Parsing

The HTML that extracted are separated using HTML tags to parse the required information.

2. Country Mapping

The location information in the reviews are mapped using its geocode to map the reviewer information to country level information.

3. Review translation

Tourists reviews that are not in English are translated to English using google translate API.

4. Keyword identification

Using POS Tag and word frequency, this process identify the keyword that relevant to each review. To identify the keywords, we make a lists of all nouns and verbs and count the occurrence of each word in the reviews. The top 1000 words then filtered manually to extract keywords that are relevant to 3 categories of tourism services: attraction, accessibilities and amenities. This process identify 68 keywords that are relevant to the aspects, 83 negative sentiment keywords, and 39 positive sentiment keywords. These 3 aspects have sub-aspects which are as follows:

- a. Attraction : Point of interests, Hospitality, and Cleanliness.
- b. Accessibilities : Infrastructures and Transportation.
- c. Amenities : Accommodation and Facilities.

Keyword Example

Table 2

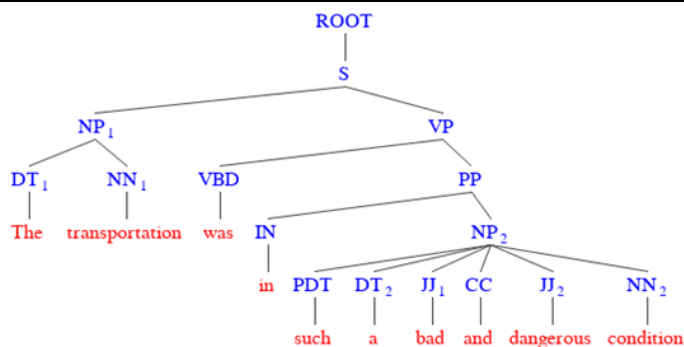
Keywords	Aspects	Sub aspects
Road	Accessibilities	Infrastructure
Street	Accessibilities	Infrastructure
Transport	Accessibilities	Transportation
Train	Accessibilities	Transportation
Coral	Attractions	Point of Interest
Island	Attractions	Point of Interest
Service	Attractions	Hospitality
Toilet	Amenities	Facilities
Hotel	Amenities	Accommodation

5. Sentence Parsing

The review text content are split into sentences and clauses. Then each sentences are parsed and decomposed into sentence parse tree structures based on English grammatical rule. For example, parse tree for the sentence *"The transportation was in such as bad and dangerous condition"* is as follows:

Sentence Parse Tree Example

Figure 3



Sentiment Analysis

This step identify the sentiments of each tourist review sentences using rule-based method based on the keyword that has been identified before. The rules for sentiment analysis are as follows:

1. Keywords for one of the aspect (attractions, accessibilities and amenities) has to be in the same clause as the sentiment keyword. For example in the sentence *"There is a lot of trash on the island, the corals are dead, and the water is dirty"* there are 3 clauses, S1, S2, and S3. In the sentence there are 3 negative attractions aspect sentiments which is identified by the keyword *"island"*, *"coral"*, and *"water"*. In S1, *"island"* keyword is next to *"trash"* keyword

which is a negative sentiment keyword. In S2, "coral" keyword is next to "dead" keyword which is a negative sentiment keyword. In S3, "water" keyword is next to "dirty" keyword.

2. If there is a negation word "not/no" in a clause, then the sentiment is reversed. For example in the sentence "The water is clean and there is no trash on the island" there are 2 positive sentiments on attraction aspect that is identified by "water" and "island" keyword. In the clause "there is no trash on the island" there is a negative sentiment keyword "trash" but there is a "no" as a negation keyword in the same clause. Therefore the clause is identified as a positive sentiment sentence on the attraction aspect.

Using the information collected from the data and sentiment sentences extracted from this process, this study create indicators of tourism destinations that can be used to analyze tourism qualities within different perspectives.

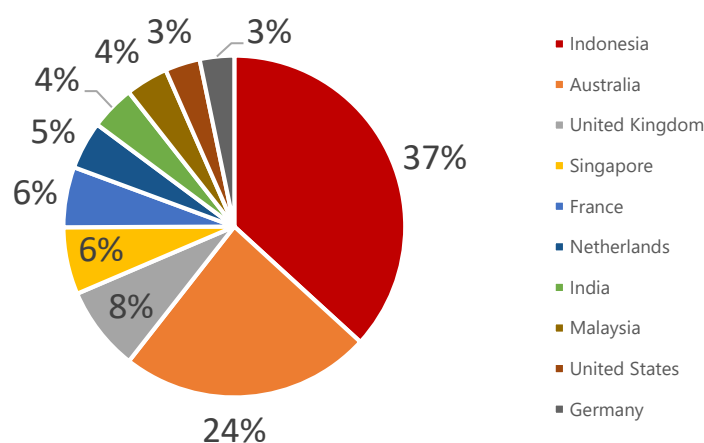
4. Result & Analysis

4.1 Tourists Country of Origins Composition

First, this study identify the composition of tourist's countries of origin from the location information that are obtained from the reviews of Indonesian tourism destination. The result is as follows:

Tourist Review Country of Origins on Indonesian Tourism Destinations (2017-2021)

Figure 4



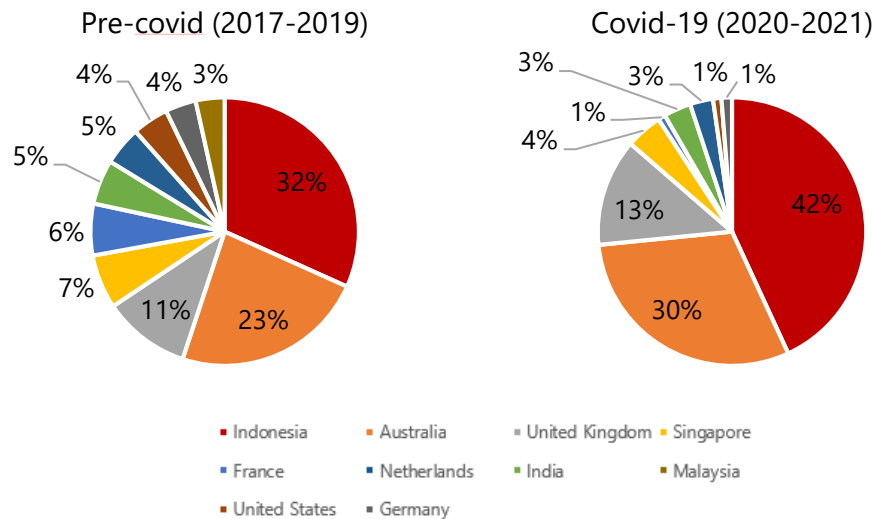
From this figure, we can see the origin of tourists that write reviews of their experience in Indonesia tourism destinations the most. Most of the reviews are from Indonesian local tourists with tourists from Australia, United Kingdom, Singapore and France from second to fifth. We can conclude that domestic tourists dominates the reviews of tourism destination the most, and Australia, UK and Singapore are the three countries that wrote reviews to the tourism destinations the most.

As we know in 2020 to 2021, international tourists are hindered to travel to Indonesia due to covid-19 pandemic that hit Indonesia and travel restrictions both

from the origin countries and Indonesian government. The reviews can be dissected by the year of visits and analyzed by pre-covid-19 (2017-2019) and covid-19 (2020-2021) years. As can be observed, there is a shift of tourist country of origins with much more of the reviews coming from local tourists (10% increase) in covid-19 era.

Tourist Review Country of Origins Pre-covid-19 and Covid-19 era on Indonesian Tourist Destinations (2017-2021)

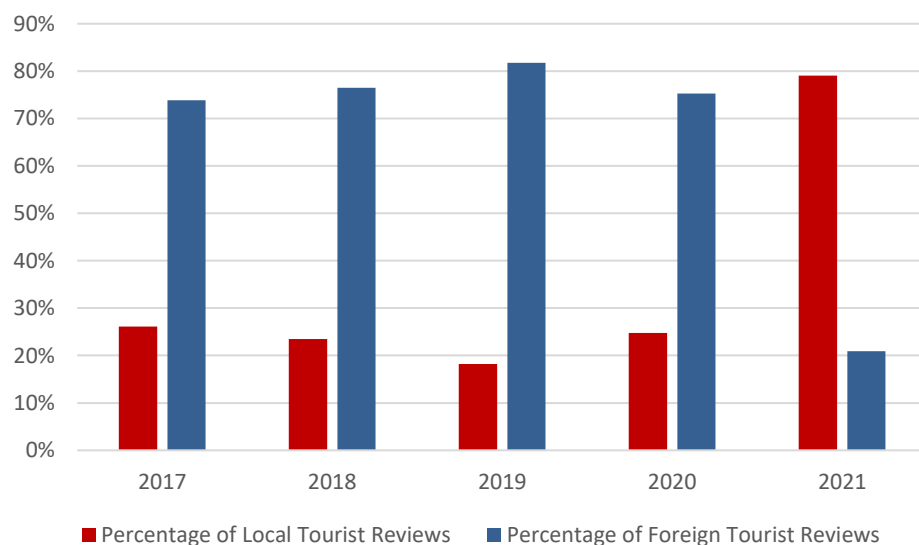
Figure 5



From the percentages of local vs foreign tourist reviews per year, in 2020 to 2021 the percentage rises significantly since there is virtually no international tourism and travel during covid-19 period especially in 2021. The percentages of local tourist reviews are declining in 2017-2019, but rises significantly in 2020 and in 2021 local tourist dominates so much more of the reviews although with low number of reviews in 2021.

Percentages of Local vs Foreign Tourists Reviews on Indonesian Tourism Destinations per Year (2017-2021)

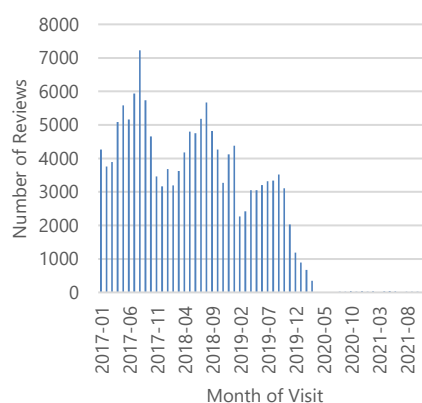
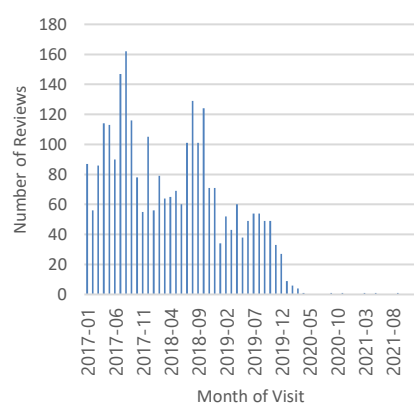
Figure 6



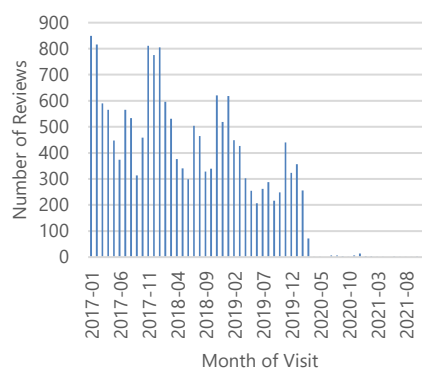
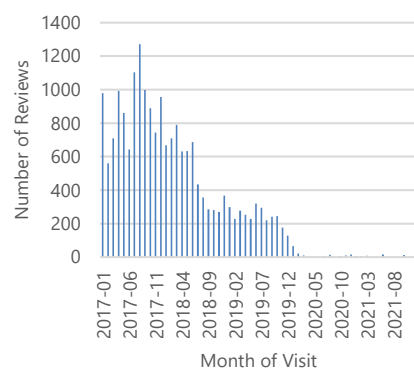
4.2 Tourism Destinations Monthly Visits

Using the visit date information that extracted from the reviews, this figures shows how the monthly visits for each tourism destinations are and when the tourism destination is more popular and have more or less visits. The effect of covid-19 pandemic to the visits of each tourism destination can be seen from the figures. The pandemic affect the visits of tourism destination as seen from the reviews with reviews reaching zero in most months in 2020-2021. Here are examples of 4 tourism destination monthly visits.

Borobudur Monthly Visits	Figure 7	Bali Monthly Visits	Figure 8
---------------------------------	----------	----------------------------	----------



Yogyakarta Monthly Visits	Figure 9	Chiang Mai Monthly Visits	Figure 10
----------------------------------	----------	----------------------------------	-----------



4.3 Tourism Destinations Rating

This study extracted the rating of tourism destinations from TripAdvisor and calculate the percentage of each of the rating from all the reviews. There are 5 level of rating (Excellent, Very Good, Average, Poor, and Terrible). Then this study ranks the tourism destination by the TripAdvisor rating to compare Indonesian tourism destinations with peer countries tourism destinations.

Tourism Destination Ranking by TripAdvisor Rating (2017-2021)

Table 3

Tourism Destinations	Rating				
	Excellent	Very good	Average	Poor	Terrible
Halong Bay	90%	7%	2%	1%	1%
Wakatobi	76%	12%	7%	3%	2%
Banyuwangi	74%	20%	4%	2%	1%
Borobudur	69%	23%	5%	1%	1%
Bromo	68%	24%	5%	2%	1%
Labuan Bajo	55%	27%	12%	3%	2%
Bangkok	51%	34%	12%	2%	1%
Ho Chi Minh City	50%	34%	12%	2%	1%
Nha Trang	50%	29%	15%	4%	3%
Danau Toba	50%	36%	10%	3%	1%
Likupang	49%	29%	20%	0%	1%
Ayutthaya	48%	36%	12%	2%	2%
Chiang Mai	46%	35%	14%	3%	2%
Yogyakarta	45%	33%	15%	4%	3%
Tanjung Kelayang	45%	42%	12%	1%	0%
Hoi An	41%	32%	19%	5%	2%
Phuket	40%	32%	18%	5%	5%
Bali	35%	41%	20%	3%	2%
Kepulauan Seribu	31%	44%	20%	3%	2%
Solo	30%	43%	22%	3%	1%
Mandalika	29%	34%	24%	7%	6%
Semarang	28%	46%	22%	3%	1%
Morotai	28%	25%	35%	10%	2%

Note: Blue-shaded cells denotes peer countries tourism destinations

By rating, the best major tourism destinations in our list is Halong Bay in Vietnam with 90% of the reviews are in excellent rating. Then in the top 10 ranking we have 5 tourism destination in Indonesia in 2nd to 6th place with excellent rating around 55% to 76%. Lastly from 7th to 9th we have 3 popular tourism destinations in Thailand and Vietnam, Bangkok, Ho Chi Minh City, Nha Trang, followed by Danau Toba¹ in 10th

¹ Danau Toba or Lake Toba is the largest lake in South East Asia and one of the deepest in the world. Located in North Sumatera, Danau Toba is one of the most promising natural attraction in Indonesia

place. Indonesian tourism destination ratings are quite competitive with tourism destinations in Thailand and Vietnam with 6 Indonesian tourism destinations in the top 10 compared to 4 from Thailand and Vietnam.

The 3A tourism quality aspects can't be observed from the rating only, since there is no additional and detailed information comes from the rating. The review sentences are rich with detailed sentiment information about the experience the reviewer has in the tourism destinations. In the next subsection this study presents the sentiment analysis results that extracted more detailed information regarding tourist experience in the destinations.

4.4 Sentiment Analysis Results

From the number of sentiment review sentences, this study shows how each tourism destinations fare in each aspects of tourism quality and how they compare with each other.

Aspect Sentiments Review Sentences (2017-2021)

Table 4

Tourism Destinations	Aspects Sentiment					
	Positive			Negative		
	Attractions	Amenities	Accessibilities	Attractions	Amenities	Accessibilities
Borobudur	1240	170	255	336	30	48
Bromo	698	96	153	296	31	94
Banyuwangi	231	8	15	54	4	74
Kepulauan Seribu	128	39	59	48	7	9
Kota Tua	183	47	44	87	14	25
Bali	48585	14738	11731	14616	3638	3428
Tanjung Lesung	22	9	9	8	1	0
Mandalika	209	51	33	76	17	11
Labuan Bajo	1506	305	324	329	92	85
Wakatobi	49	15	13	8	2	1
Morotai	6	0	4	8	2	0
Danau Toba	116	39	35	13	1	6
Tanjung Kelayang	44	12	10	8	2	1
Yogyakarta	3412	640	784	1090	154	367
Solo	336	98	91	63	26	23
Semarang	432	104	109	130	37	47
Batam	887	329	194	269	80	69
Likupang	20	10	8	13	4	1

Aspect Sentiments Review Sentences (2017-2021)						Table 4
Chiang Mai	1939	624	336	1179	392	386
Bangkok	2612	629	831	2406	860	1204
Phuket	2124	699	589	2109	822	925
Ayutthaya	368	90	91	306	97	100
Ho Chi Minh City	1724	362	307	1263	377	347
Halong Bay	1190	635	585	797	410	291
Nha Trang	1085	218	179	867	154	172
Hoi An	1947	444	286	1346	343	325

Note: Blue-shaded cells denotes peer countries tourism destinations

Indonesian popular tourism destinations such as Bali and Yogyakarta has more reviews compared to Thailand and Vietnam tourism destinations, while Indonesian less popular destinations have significantly little number of reviews. Generally there are more positive than negative review sentences across all tourism destinations. TripAdvisor reviews tend to be more positive than negative, suggesting that reviewers like to share about their travelling experiences and they not only use the platform to complaint about the bad experiences they may have.

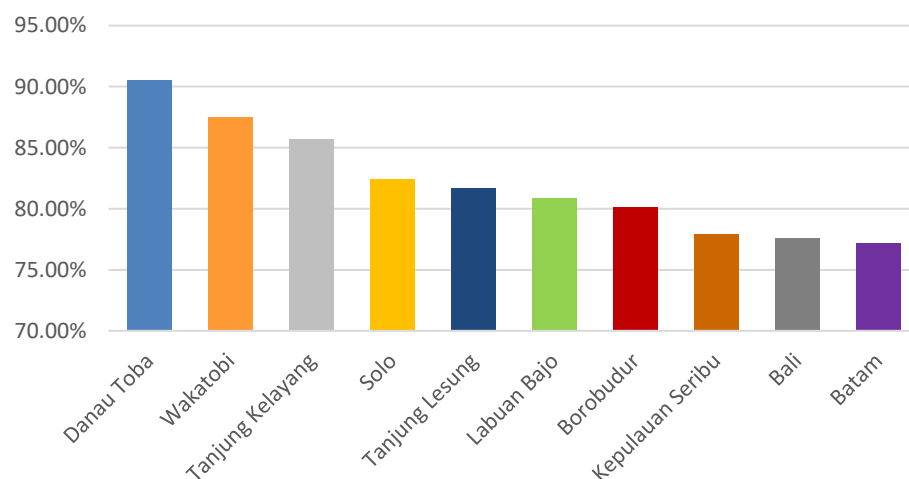
From the number of review sentences, the tourism destinations are ranked by the shares of its positive sentiments from all sentiment review sentences. Tourism Destinations like Danau Toba, Wakatobi², and Tanjung Kelayang³ has the highest shares of positive sentiments compared to other tourism destinations. From this perspective, Indonesia's tourism destinations ranks in all the Top 10 destination from the 26 tourism destinations.

² Wakatobi Regency is a group of ca. 150 islands forming an administrative regency located in Southeast Sulawesi, Indonesia. The four largest islands are Wangi-wangi, Kaledupa, Tomia and Binongko. The Wakatobi Islands are a part of the Coral Triangle, which contains one of the richest marine biodiversity on earth.

³ Tanjung Kelayang is a Special Economic Area that is developed for its beach tourism. It's located in Belitung Island and famous for blue and clear seawater and also the soft, white, and beautiful sand in the shorelines.

Top 10 Tourism Destinations by Shares of Positive Sentiment in All Aspects (2017-2021)

Figure 11

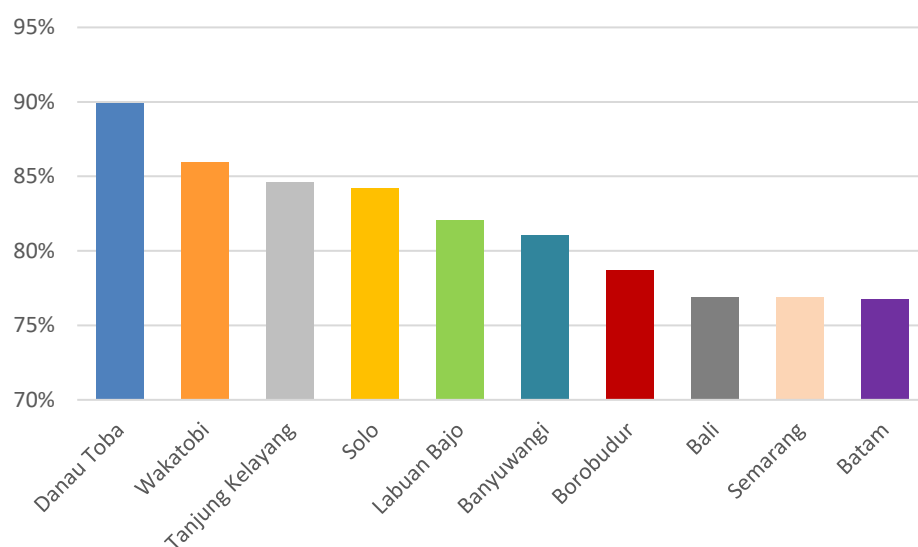


This ranking can be used to give recommendation to the government of which tourism destinations that are better priority to promote more intensely. Danau Toba, Wakatobi and Tanjung Kelayang are three tourism destinations that doesn't have that many review sentiment sentences and can be considered underrated. With better promotional strategies, we can expect improvement of tourist visits to these three tourism destinations after the covid-19 pandemic ends.

The ranking can also be decomposed for each qualities aspects.

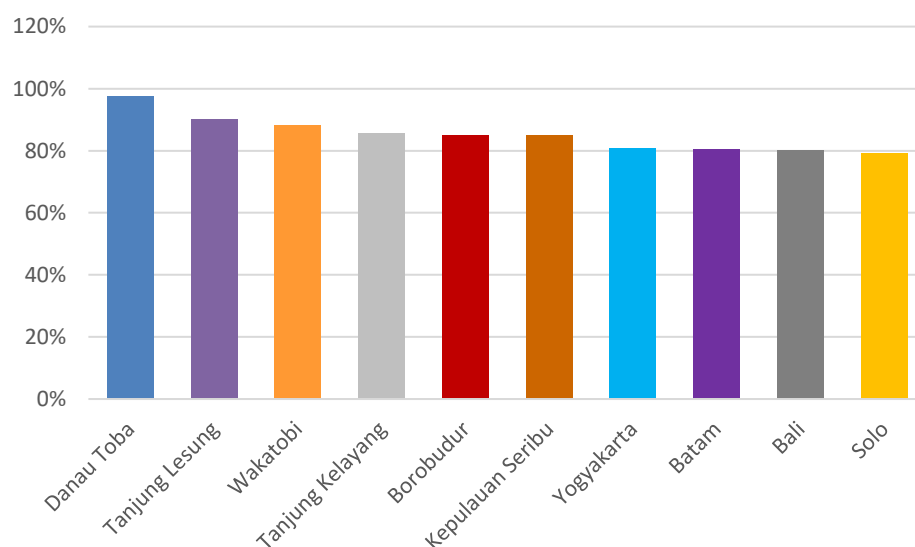
Top 10 Tourism Destinations by Shares of Positive Sentiment in Attraction Aspects (2017-2021)

Figure 12



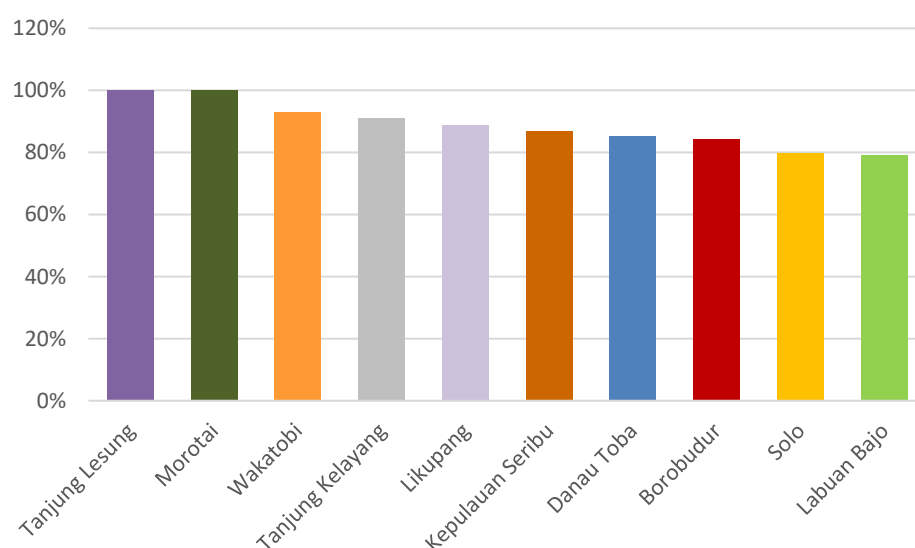
Top 10 Tourism Destinations by Shares of Positive Sentiment in Amenities Aspect (2017-2021)

Figure 13



Top 10 Tourism Destinations by Shares of Positive Sentiment in Accessibilities Aspect (2017-2021)

Figure 14



This table shows the granular details of review sentences that we have obtained using this methodology. We are able to extract positive and negative sentiments sentences that we can use to understand more about the main aspects for promotion and improvement for each of the tourism destinations. The positive sentiments can be used to formulate policy to strengthen promotional aspects of tourism destinations, while the negative sentiments can be used to recommend aspects of tourism quality that needs to be improved.

Tourism Destinations Sentiment Examples Top 10 Tourism
Destination by Positive Review Shares

Table 5

Tourism Destinations	Sentiment Reviews
Danau Toba	The lake and its surrounding hills are beautiful just to watch.
	What is lacking is the infrastructure to make this place a real tourist spot.
Wakatobi	The care was always excellent, diving places could be gotten in the short term.
	Ferry schedule/service is a bit tough due to changing weather and local conditions.
Tanjung Kelayang	This beach is quite managed neatly, because the beach is clean, on the beach there are also many boats that can be rented for Hopping Island
	Unfortunately the boat rental is quite expensive 500rb / boat with the contents of max 7 people
Solo	Very interesting collection of batik textile as a piece of Indonesian history. Excellent helpful very kind guide tour we had only for our family.
	Closed for unadvertised lunch between 11.30am and 1pm so we couldn't even get into the museum. Staff were pretty unfriendly so we left.
Tanjung Lesung	The beach club provides well-maintained bathrooms and prayer facilities. The beach is also quite clean and well-maintained
	After getting very disappointed, i kept my faith and saw the board which listed the prices. i was shocked. An ATV rental for 20 min (max 2 ppl) for 300,000 Rp. and Jet-ski for 120,000 Rp.
Labuan Bajo	Very very beautiful beach. We also had a very nice snorkeling. Would really love to be back here again!
	If I knew road was this steep and half paved bumpy, I wouldn't go. Not recommend to older adult.
Borobudur	Went in before sunrise with a guide who helped find the picturesque spots. The temple is amazing, historic and beautiful.
	Ticket prices are ridiculous high for foreigners. Locals paid only IDR 40,000 whereby for foreigners the price is 10 times higher. It is ridiculous. Nothing much to view.
Kepulauan Seribu	This was a visit to the volunteer of Earth's Day with LIPI on the Thousand Islands Pari Island...The sunset is pink and beautiful ...
	Visited island with my wife. After 10 minutes we arrived my wife started to cry from shock. The island is full of rubbish, dirty, water is dirty too.
Bali	We had super good time. The staffs were nice and helpful, they arranged 3 round beach benches for our group. The beach clean, water nice and warm, music nice. Foods also good in here. Everything.

Tourism Destinations Sentiment Examples Top 10 Tourism
Destination by Positive Review Shares

Table 5

	Boat was overcrowded and life jackets not in place. Sure the boat got there but if there was a problem unnecessary casualties would have occurred.
Batam	It's a Chinese Buddhist temple very near to city centre. It is a very silence place & a suitable place for worship.
	Very bad experience arriving at this Terminal. Queue was long, custom was rude and clearance was slow.

Notes: Green shaded cells are positive sentiment reviews. Red shaded cells are negative sentiment reviews

From review sentiment sentences, we can conclude that Indonesia tourism destinations qualities in each aspects are as follows:

1. Attractions : While Indonesia has strong natural and cultural attractions that praised by most of the review sentiments, the cleanliness of its attractions hasn't maintained well in some of the tourism destinations.
2. Amenities : Most of the reviews express good review of clean facilities, but there are still complaints in some of the places with dirty toilets and facilities are the most frequent complaints. Other frequent complaints are also regarding the cost and comfort of transportation within the tourism destinations.
3. Accessibilities: The access and cost of travel to the tourism destination is one of the most frequent complaints in many of the less popular destinations. The accessibilities has to be improved for many of the less popular tourism destinations for more tourists to become recurring visitors of the tourism destinations.

Generally our recommendations to improve the quality of most Indonesia tourism destinations from the result of this study are as follows:

1. Promote its natural and cultural attraction aspect more intensely.
2. Develop a framework to maintain consistent and thorough cleanliness in the tourism destinations locations.
3. Carefully take attention to the transportation services to and within the tourism destinations, especially regarding the cost and quality.

5. Conclusion & Future Works

5.1 Conclusion

TripAdvisor API is a powerful tools to extract large amount of reviews that can be utilized to discover strengths and weaknesses of tourism destinations. Using TripAdvisor API we extracted 604,395 reviews of 26 tourism destinations in Indonesia, Thailand and Vietnam from 2004-2021 and analyze it using sentiment analysis to gather information about how the quality of each tourism destination in 3A (Attractions, Amenities, Accessibilities) tourism quality aspects. We also compare

Indonesia tourism destinations with peer countries (Thailand and Vietnam) tourism destinations using the rating that we obtained from TripAdvisor and the sentiment analysis results.

From the TripAdvisor rating collected, Indonesia have 6 tourism destinations in the top 10 with shares of excellent rating ranges from 50% to 76% compared to 4 peer countries tourism destinations ranks in the top 10 with shares of excellent rating from ranges from 50% to 90%. From sentiment perspectives, Indonesian tourism destination generally have better positive sentiment shares compared to peer countries tourism destination.

The top 3 tourism destinations by positive sentiment shares are Danau Toba, Wakatobi, and Tanjung Kelayang. From the review sentences extracted we can formulate better strategies promotion to improve the 3A aspect of each tourism destination qualities.

Our recommendations for most Indonesia tourism destinations from the result of this study are as follows:

1. Promote its natural and cultural attraction aspect more intensely.
2. Develop a framework to maintain consistent and thorough cleanliness in the tourism destinations locations and the facilities around it.
3. Carefully take attention to the transportation services to and within the tourism destinations, especially regarding the cost, quality, and comfort.

5.2 Future Works

These are some room for improvements and future works that may be done.

There are other travel platform reviews e.g Yelp, Google Places, Booking.com, etc that may have useful similar data. Although TripAdvisor is the most popular tourist review platform, there may be other interesting findings if we perform similar methodology to data from other travel platforms.

More sentiment reviews can be extracted with expansion of aspects and sentiment keywords list. The study has used 1000 of the most frequent English nouns and verbs, but there may have some nouns and verbs that can be used to expand the scope of the aspects and sentiment keywords.

Similar sentiment analysis methodology can be done for accommodation reviews data especially those around tourism destinations. This study will be helpful to understand the qualities of accommodation which is part of the second tourism quality aspect (Amenities) in a more detailed manner.

Sharia tourism is also one of the interesting perspective of tourism that can be looked more closely. As the biggest Muslim country in the world, Indonesia has the potential to capture the majority of Sharia tourist market share. Similar study can be done using sharia tourism specific aspects keyword such as praying room, mosque, halal restaurant, etc.

References

- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24.
- Chua, A. Y., & Banerjee, S. (2013). Reliability of reviews on the Internet: The case of TripAdvisor. In *World Congress on Engineering & Computer Science: International Conference on Internet and Multimedia Technologies* (pp. 453-457). York.
- Ekman, P. (1992). Facial expressions of emotion: New findings, new questions.
- Hearst, Marti. "What is text mining." *SIMS, UC Berkeley* 5 (2003).
- Inskip, E. (1991). *Tourism planning: An integrated and sustainable development approach*. John Wiley & Sons.
- Lo, I. S., McKercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism management*, 32(4), 725-731.
- Mason, P. (2020). *Tourism impacts, planning and management*. Routledge.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- Middleton, V. T., & Clarke, J. R. (2012). *Marketing in travel and tourism*. Routledge.
- Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of travel research*, 46(1), 35-45.
- Park, S. B., Ok, C. M., & Chae, B. K. (2016). Using Twitter data for cruise tourism marketing and research. *Journal of Travel & Tourism Marketing*, 33(6), 885-898.
- Thaha, A. R., & Aziz, F. (2020). Text Mining on Tourism Destinations in Bandung Raya (Case Study: Tangkuban Perahu and Kawah Putih). *Secretary and Business Administration Journal*, 4(2), 146-156.
- Yoeti, Oka A. (1997). *Planning and Development of Tourism*, PT. Pradnya Paramita: Jakarta.

Sentiment Analysis

of Tourist Reviews from Online Travel Forum

for Improving Indonesia Tourism Sector

**Anggraini Widjanarti, Muhammad Abdul Jabbar, Arinda
Dwi Okfantia, Alvin Andhika Zulen.**

IFC-Bank of Italy: Data science in central banking, 14-17 February 2022

*The views expressed here are those of the authors and do not necessarily reflect the views of Bank Indonesia



OUTLINE

2



BACKGROUND AND GOALS



FRAMEWORK



METHODOLOGY



RESULTS AND ANALYSIS



CONCLUSION AND FUTURE WORKS



BACKGROUND AND GOALS

3

Background

1. Tourism is one of strategic priority sector in Indonesia with big multiplier effects to the economy.
2. Discovering big picture and granular details of tourism reviews in popular tourism destinations
3. Untapped Big Data sources (TripAdvisor reviews) that can be analyzed using text mining

Goals

- We extract tourist reviews in Indonesia's and peer countries major tourism destinations to analyze its 3A's tourism qualities aspects (Accessibilities, Accommodation, Amenities).
- Collect positive reviews to strengthen the promotional aspects of tourisms, and negative reviews to identify recommendation for improvement in tourism destinations.

Literature Study

1. Tourism Development Aspects:

Inskeep, E. (1991). Tourism planning: An integrated and sustainable development approach: The study mainly talks about the aspects of tourism development.

"Tourism development involves components regarding tourism planning approaches, tourism attractions, accommodation, tourism facility and other tourism services such as transportation, infrastructure, and tourism management institutions."

2. TripAdvisor Reviews Credibility:

Chua, A. Y., & Banerjee, S. (2013). Reliability of reviews on the Internet: The case of TripAdvisor: The paper studies the reliability of TripAdvisor review by using Social Network Analysis of hotel reviewers in Singapore.

"On establishing review baselines, the inter-reviewer and intra-reviewer reliability of contributions was studied using reviewer-centric and review-centric approaches respectively. Results suggest that reviews in TripAdvisor could be largely reliable."

3. Text Mining on Tourism Data

Thaha, A. R., & Aziz, F. (2020). Text Mining on Tourism Destinations in Bandung Raya (Case Study: Tangkuban Perahu and Kawah Putih). Similar sentiment analysis study using TripAdvisor review data on Indonesia tourism destinations.

"Research findings illustrate that text mining can be applied to travel reviews for tourism destinations in Indonesia."



FRAMEWORK & SCOPE

4

To understand Indonesia's tourism competitive advantages we analyze sentiments to 3 aspects of tourism qualities (Accessibilities, attractions, and amenities) from 18 popular tourism destinations from Indonesia and 8 popular tourism destinations from Vietnam and Thailand.

Framework

1



Review extraction using TripAdvisor API

2



Data pre-processing to structure the data and enrich the information in the data

3



Sentiment analysis using text mining to extract sentiment sentences from the reviews

Scope

Tourism Quality Aspects

Accessibilities

1. Infrastructure
2. Transportation

Attractions:

1. Point of interests
2. Hospitality
3. Cleanliness

Ammenities:

1. Accommodation
2. Facilities

Tourism Destinations

Indonesia:

1. Borobudur
2. Bromo
3. Banyuwangi
4. Kepulauan Seribu
5. Kota Tua
6. Yogyakarta
7. Tanjung Lesung
8. Mandalika
9. Labuan Bajo
10. Wakatobi
11. Morotai
12. Danau Toba
13. Tanjung Kelayang
14. Bali
15. Solo
16. Semarang
17. Batam
18. Likupang

Peer Countries:

1. Chiang Mai
2. Bangkok
3. Phuket
4. Ayutthaya
5. Ho Chi Minh City
6. Halong Bay
7. Nha Trang
8. Hoi An



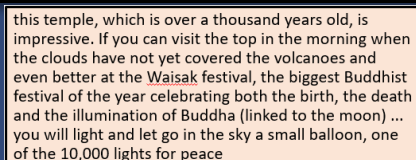
Country Mapping



Name	Dina K
Tourist Origins	Izhevsk

Name	Dina K
Tourist Origins	Izhevsk
Country of Origins	Rusia

Sentence Split & Parsing



But/CC what/WP we/PRP found/VBD was/VBD sad/JJ sad/JJ
sad/JJ . There/EX is/VBZ a/DT lot/NN of/IN trash/NN on/IN
the/DT island/NN , the/DT corals/NN are/VBP dead/JJ, and/CC
the/DT water/NN is/VBZ dirty/JJ.

Keyword	Aspek	Sub Aspek	Keyword
Road	Accessibilities	Infrastructure	Dirty
Street	Accessibilities	Infrastructure	Bad
Transport	Accessibilities	Transportation	Horrible
Train	Accessibilities	Transportation	Poor
Coral	Attractions	Point of Interest	Pricey
Island	Attractions	Point of Interest	Smell
Service	Attractions	Hospitality	Impolite
Fee	Attractions	Price	Lack
Trash	Attractions	Cleanlines	Rip off, etc

We booked a local tour and departed to nowhere, expecting to find a paradise looking island in the middle of nowhere. But what we found was sad sad sad. There is a lot of trash on the island, the corals are dead, and the water is dirty. Still, the people were very welcoming and our tour guide tried hard to make our stay enjoyable. We did some snorkelling, but it just reminded us about how much harm we have caused to the environment.

Sentence Splitting

1. We booked a local tour and departed to nowhere, expecting to find a paradise looking island in the middle of nowhere.
2. But what we found was sad sad sad.
3. There is a lot of trash on the island, the corals are dead, and the water is dirty.
4. Still, the people were very welcoming and our tour guide tried hard to make our stay enjoyable.
5. We did some snorkelling, but it just reminded us about how much harm we have caused to the environment.



METHODOLOGY – SENTIMENT EXTRACTION

6

Sentiments are extracted from the review sentences data using rule-based methodology. The rules for the extraction are as follows:

- 1 Keywords for one of the attractions, accessibilities and amenities aspect has to be in the same clause as the sentiment keyword.
- 2 If there is a no/not word in the clause, the sentiment is reversed (positive to negative, vice versa)

Example

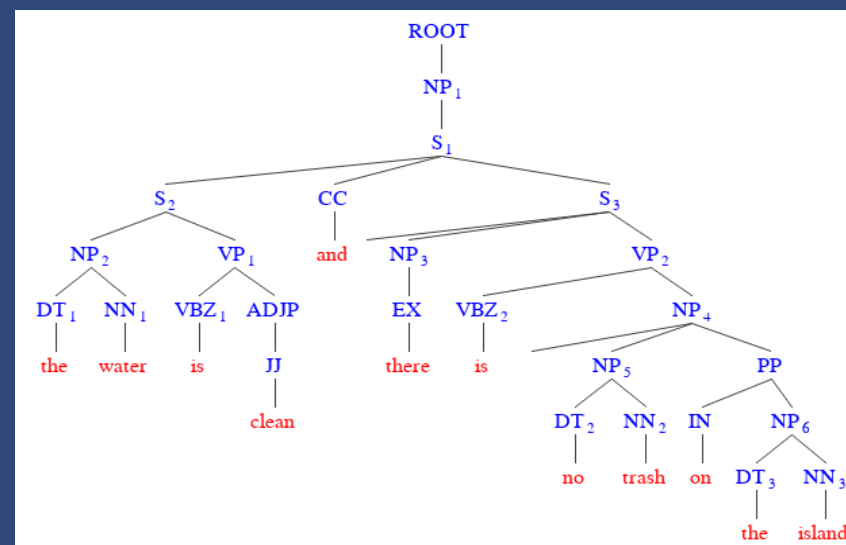
“The water is clean and there is no trash on the island”

There are 2 clauses in the review sentence:

S1: The water is clean

S2: There is no trash on the island

1. S1 is a positive sentiment sentence because there is positive sentiment keyword 'clean' and attraction keyword 'water'
2. S2 is a positive sentiment sentence because there is negation word 'no' with negative sentiment keyword 'trash', and attraction keyword 'island'



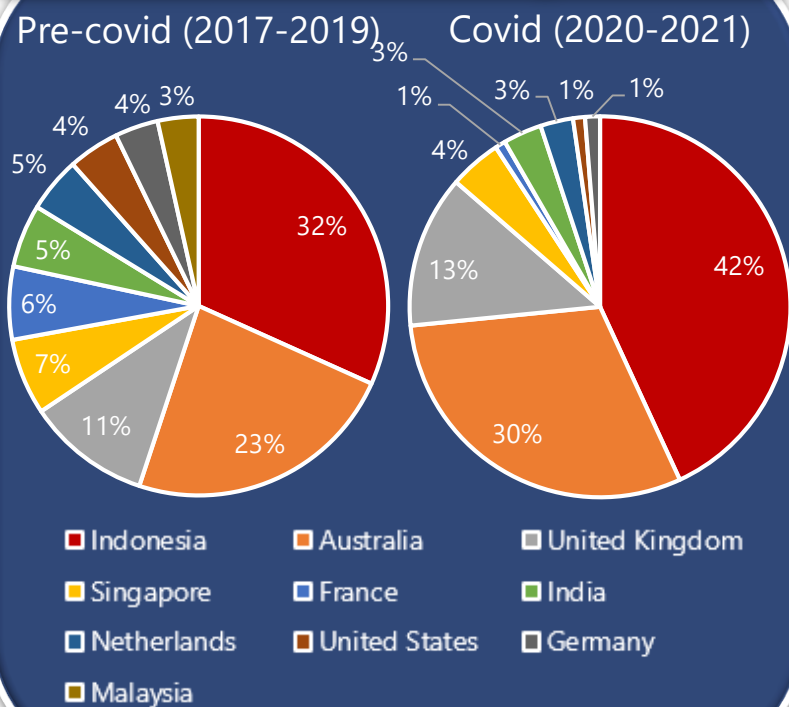


RESULT – GENERAL OVERVIEW

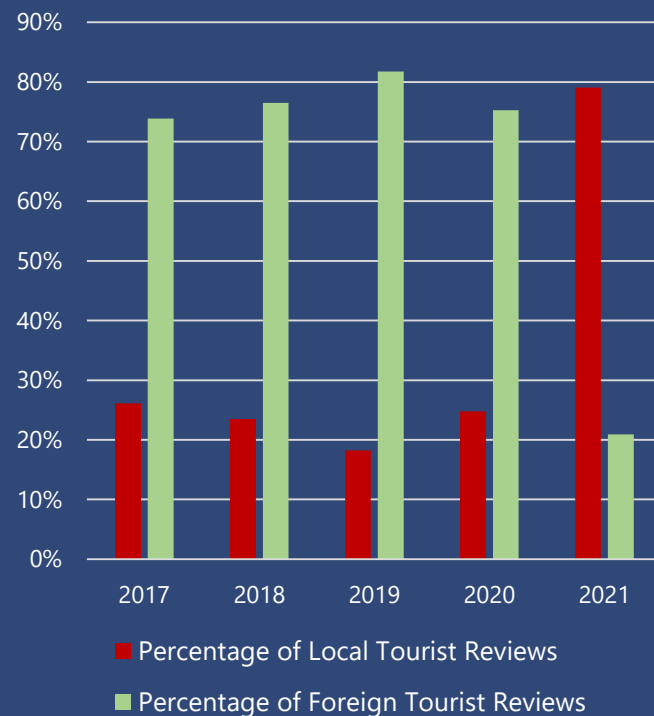
7

- As the covid-19 pandemic causes travel restrictions around the world, we can see the shift of tourists composition in Indonesia travel destinations. It's also shown in the number of reviews in TripAdvisor. On Covid era, local tourists dominates the review compared to foreign tourists.
- Halong Bay is the top tourists destination by TripAdvisor rating followed by 5 tourism destination from Indonesia.

Top 10 Tourist Reviews Countries of Origin*



Percentages of Local vs Foreign Tourist Reviews*



Top 10 Tourist Destinations by Rating 2017-2021

Tourist Destination	Rating				
	Excellent	Very Good	Average	Poor	Terrible
Halong Bay	90%	7%	2%	1%	1%
Wakatobi	76%	12%	7%	3%	2%
Banyuwangi	74%	20%	4%	2%	1%
Borobudur	69%	23%	5%	1%	1%
Bromo	68%	24%	5%	2%	1%
Labuan Bajo	55%	27%	12%	3%	2%
Bangkok	51%	34%	12%	2%	1%
Ho Chi Minh City	50%	34%	12%	2%	1%
Nha Trang	50%	29%	15%	4%	3%
Danau Toba	50%	36%	10%	3%	1%

*Indonesia tourism destinations only, 2017-2021 period

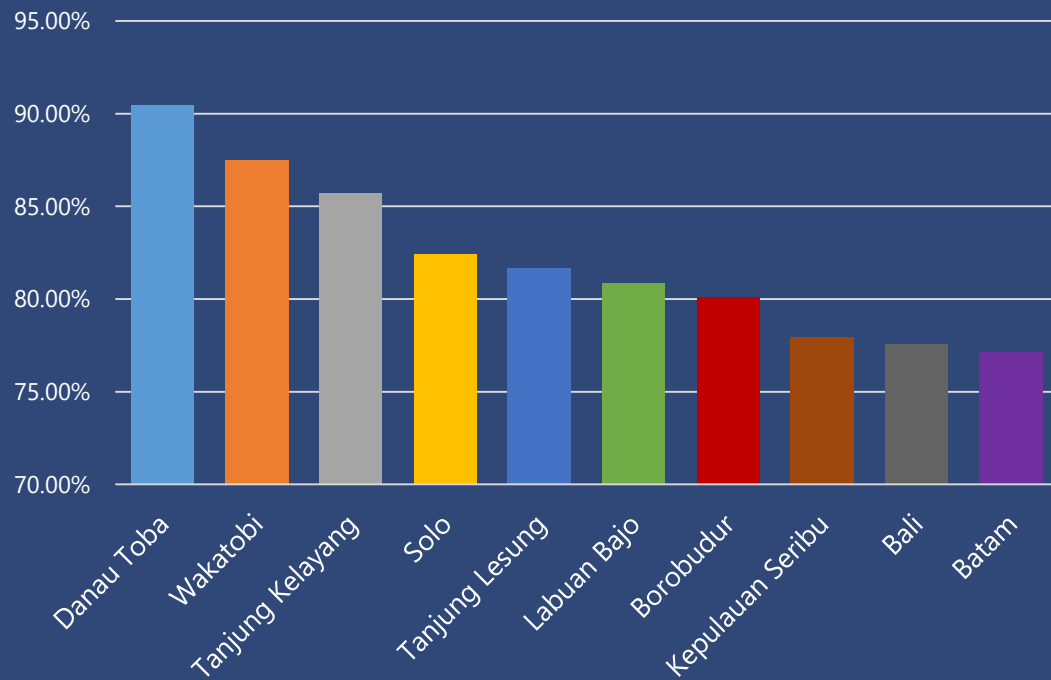


RESULT – SENTIMENT ANALYSIS

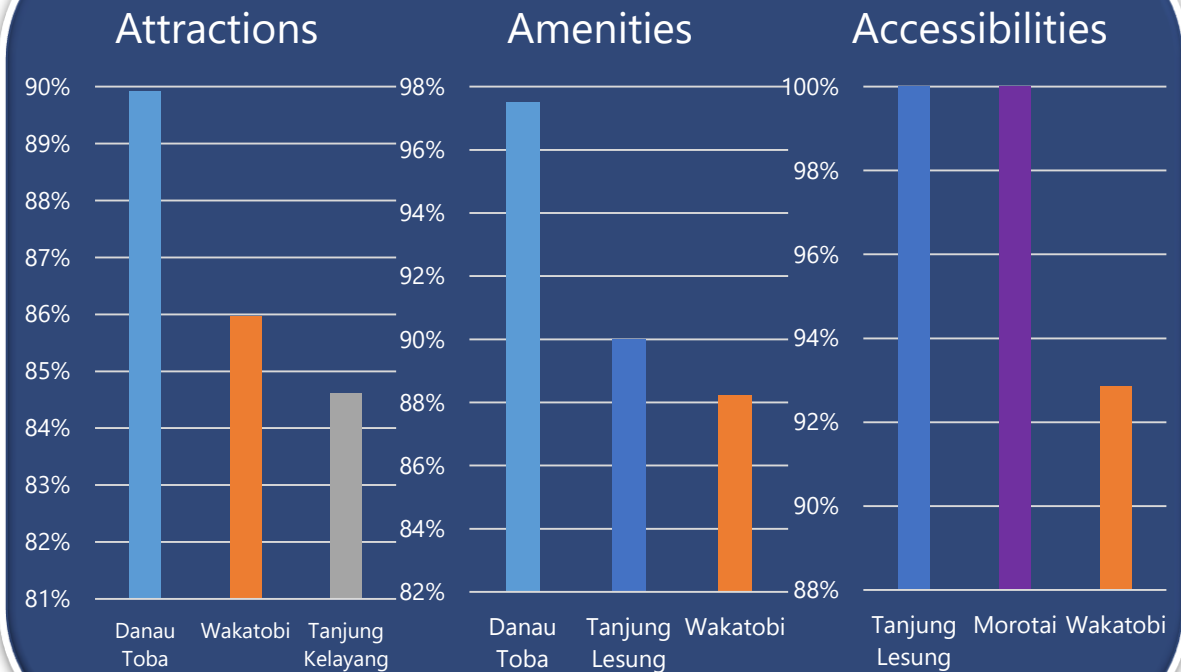
8

From the extracted sentiments we can rank the tourism destinations by its shares of positive vs negative sentiments. We can also further scrutinize the data and rank by its 3A aspects. From sentiments perspectives, Indonesia's tourism destination ranks in all the top 10.

Top 10 Tourism Destination by Shares of Positive Sentiments 2017-2021



Top 3 Tourism Destination in 3A Aspects by Shares of Positive Sentiments 2017-2021





RESULT – SENTIMENT SENTENCES

9

From the review sentences we can see the strengths and weaknesses of each tourism destination and use them to analyze aspects to promote and to improve about for each of the tourism destinations.

Danau Toba Sentiment Sentences



- 1. this area is an absolute beauty.
- 2. The lake and its surrounding hills are beautiful just to watch.

- 1. What is lacking is the infrastructure to make this place a real tourist spot.
- 2. Also, cleanliness must be improved.

Wakatobi Sentiment Sentences



The care was always excellent, diving places could be gotten in the short term.

Ferry schedule/service is a bit tough due to changing weather and local conditions.

Tanjung Kelayang Sentiment Sentences



This beach is quite managed neatly, because the beach is clean, on the beach there are also many boats that can be rented for Hopping Island

unfortunately the boat rental is quite expensive 500rb / boat with the contents of max 7 people



Conclusion

1. TripAdvisor API is a powerful tools to extract large amount of reviews that can be utilized to discover strengths and weaknesses of tourism destinations.
2. Indonesian tourism destination generally have better positive sentiment shares compared to peer countries tourism destination.
3. The top 3 tourism destinations by positive sentiment shares are Danau Toba, Wakatobi, and Tanjung Kelayang. From the reviews extracted we can formulate better strategies to promote and improve the 3A aspect of each tourism destination qualities.

Future Works

1. Exploration of other travel platform reviews e.g Yelp, Google Places, Booking.com, etc.
2. Expansion of aspects and sentiment keywords.
3. Accommodation around tourism destinations review extractions.
4. Sharia tourism specific review extractions.

A world map is shown in the background, colored with a gradient from light blue at the top to dark red at the bottom. The text "Thank you!" is centered over the map.

Thank you!