Murga J. et al. (2020) A Sentiment Analysis Software Framework for the Support of Business Information Architecture in the Tourist Sector. In: Hameurlain A. et al. (eds) Transactions on Large-Scale Data- and Knowledge-Centered Systems XLV. Lecture Notes in Computer Science, vol 12390. Springer, Berlin, Heidelberg.

DOI: https://doi.org/10.1007/978-3-662-62308-4_8

A Sentiment Analysis Software Framework for the support of Business information architecture in the tourist sector

Javier Murga[1], Gianpierre Zapata[2], Heyul Chavez[3], Carlos Raymundo[4], Luis Rivera[5], Francisco Domínguez[5], Javier M. Moguerza[5], and José María Álvarez[6]

[1] Ingeniería de Software, Universidad Peruana de Ciencias Aplicadas, Lima, Perú
u201111811@upc.edu.pe
[2] Ingeniería de Sistemas de Información, Universidad Peruana de Ciencias Aplicadas, Lima, Perú
u201214895@upc.edu.pe
[3] Ingeniería de Telecomunicaciones y Redes, Universidad Peruana de Ciencias Aplicadas, Lima, Perú
u812426@upc.edu.pe
[4] Dirección de Investigación, Universidad Peruana de Ciencias Aplicadas, Lima, Perú
carlos.raymundo@upc.edu.pe
[5] Escuela Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Mostoles, Madrid, España
lm.rivera@alumnos.urjc.es, {francisco.dominguez,javier.moguerza}@urjc.es
[6] Department of Computer Science and Engineering, Universidad Carlos III, Madrid, Spain
joalvare@inf.uc3m.es

Abstract. In recent years, the increased use of digital tools within the Peruvian tourism industry has created a corresponding increase in revenues. However, both factors have caused increased competition in the sector that in turn puts pressure on small and medium enterprises' (SME) revenues and profitability. This study aims to apply neural network based sentiment analysis on social networks to generate a new information search channel that provides a global understanding of user trends and preferences in the tourism sector. A working data-analysis framework will be developed and integrated with tools from the cloud to allow a visual assessment of high probability outcomes based on historical data, to help SMEs estimate the number of tourists arriving and places they want to visit, so that they can generate desirable travel packages in advance, reduce logistics costs, increase sales, and ultimately improve both quality and precision of customer service.

Keywords: Sentiment analysis · Framework · Predictive · Tourism · Cloud computing

1 Introduction

The tourism industry is an important sector of the world economy. In 2018, global tourism increased 6% according to the United Nations World Tourism

Organization (UNWTO), and it is expected to in-crease by 3% to 4% in 2019. Despite the political and economic conflicts that continually arise around the world, global tourism metrics continues to rise. According to the Ministry of International Commerce and Tourism (MINCETUR in Spanish), in 2018, approximately 4.4 million foreign tourists visited Peru, 9.6% more than 2017, and generated US $ 4.9 billion of revenue [1].

SMEs constitute the majority of tourism service providers nationwide, however, they lack the infrastructure and resources to provide products and services to a greater number of people, since in order to be profitable they need to have a minimum of packages purchased. For this reason, if they do not manage to have this minimum of packages purchased, they need to outsource certain services to a larger tourism company and reduce their profit margins. In addition to reduced profits, there is considerable risk, as separating some of these services requires a non-refundable pre-deposit.

With this study, a framework is proposed that will predict the number of tourists who will visit an area and the packages they want to take, through a long-term sentiment analysis and a short-term, real-time analysis of Twitter posts. By applying sentiment analysis to these comments and posts, the system estimates a general level of satisfaction with the destinations visited and can thus estimate an increase or decrease in tourists during specific time periods. These statistics will allow SMEs to significantly reduce the risk of not reaching the minimum of packages purchased, and will even allow them to start building personalized tour packages as they will meet the interests of tourists, and can improve their profit margins.

2 Literature Review

2.1 Analysis of different languages

In the relevant research of language analysis, [3] developed a sentiment analysis system for the two most used languages in Malaysia, English and Malay, focusing on the lexicon, as the majority of the published research on sentiment analysis has concentrated on the vocabularies of the English lexicon. However, [4] presented a language-independent sentiment analysis model, with the domain based on n-grams of characters to improve classifier performance using the surrounding context. The results confirmed that this approach of integrating the surrounding context was more effective for data sets of different languages and domains. This suggests that a model based on n-grams of characters for data sets of multiple domains and languages is effective. Thus, a simple all-in-one classifier, that uses a mix of labeled data in multiple languages (or domains) to train a model of sentiment classification, can compete with more sophisticated domain or language adaptation techniques. On the other hand, [5] presents an innovative solution that considers space and temporal dimensions, using automatic geolocation techniques, for sentiment analysis of users that have a sense of belonging to a group. Geolocation is language independent and does not make previous assumptions about the users.

In the case of the articles mentioned above, the outcomes show that the sentiment analysis can be reliable. However, the proposed methods in [3], [4] and [5] produce average accuracy due to the use of the slang, abbreviated words and dialects widely used in social networks and thus difficult to decipher.

2.2 Sentiment Analysis in the Tourism Industry

In relation to sentiment analysis in the tourism industry, [6] and [7] designed a model to analyze hotel customers' comments, [8] analyzed the flight experience of airline passengers from their social network comments, and [9] applied different sentiment analysis approaches for tourism in general, reviewing and evaluating them in terms of data sets used and performance relative to key evaluation metrics.

Nevertheless, most of the available hotel review or flight experience text data sets lack labels. As they represent feelings, attitudes and opinions that are commonly full of idiomatic expressions, onomatopoeias, homophones, phonemes, alliterations and acronyms, they are difficult to decipher and require a large amount of work to pre-process [6], [7], [8]. In particular, [8] uses sentiment analysis techniques to analyze negative, neutral and positive feelings in relation to the top ten airlines in the United States. And [9] outlines future research in tourism analysis as part of an expansive, Big Data approach.

2.3 Social Network

Regarding social networks, [10] developed a Sentiment Analysis Engine (SAE) that estimates the sentiment of users in terms of positive, negative or neutral polarity. Their SAE is based on the classification of an automatic text learning model, trained by real data flows deriving from different social network platforms that specialize in user opinion (for example, TripAdvisor). Monitoring and sentiment classification are then carried out on the continuously extracted comments from publicly available social networks such as Facebook, Twitter and Instagram, a procedure that [11] performs as well. In a specific case, [12] presents a model for analyzing the impact of a brand by fusing real data collected from Twitter over a 14-month period, and also analyzes the revisions that covers the existing methods and approaches in the sentiment analysis. In a more general case, [13] suggests a framework consisting of analysis modules and linguistic resources where two main analysis modules are run by a classification algorithm that automatically assigns class appropriate labels of intent and sentiment for a given text. However, [14] demonstrates that addressing negation can improve the final system and thus developed an unsupervised polarity classification system, based on the integration of external knowledge. To evaluate this influence, a group of tweets were first analyzed by their suggested unsupervised polarity classification system to detect negation, and then under a sentiment analysis that considered their detected negation, and a control that didn't. As seen above, traditional sentiment analysis emphasizes the classification of web comments in positive, neutral and negative categories. However, [15] goes beyond classification of sentiments by focusing on techniques that can detect the specific topics that correspond to positive and negative opinions. Combining these techniques can help understand the general reach of sentiment as well as sentiment drivers. Contrary to the articles previously mentioned, [16] analyzes the textual content as well as the visual one. As the old saying goes, "a picture is worth a thousand words", and the image tweet is a great example of a multimodal sentiment.

In conclusion, each article reviewed here has a different approach in analyzing social network sentiment, as they attack the problem from their individual perspective. For example, [16] focuses on the sentiment analysis based on visual and multimedia information. The results obtained in [14] reveal that the analysis of negation can greatly improve the accuracy of the final system. And [11] concludes that information extraction techniques based on Twitter allow for the collection of direct answers from a target public, and therefore provide a valuable understanding of public sentiment to predict an overall opinion of a specific product. In order to train its classification model, [13] suggests the linguistic resources of corpus and lexicon. Corpus consists of a collection of texts manually labeled with the appropriate classes of intention and sentiment. Lexicon consists of general terms of opinions and clusters of words that help to identify the intentionality and the sentiment. This later process requires manual entry of a large quantity of information and is therefore quite complicated and time-intensive.

2.4 Types of Neural Networks

In recent years, deep artificial neural networks, including recurrents, have won numerous pattern recognition and machine learning competitions. [17] summarizes succinctly the significant work of the last millennium. Shallow and deep learners are distinguished by the depth of their credit allocation routes, which are the chains of possibly learnable, causal links between actions and effects. Deep supervised learning, non-supervised learning, reinforcement learning, evolutive calculation, and indirect search of short programs that codify big and deep networks are reviewed. [18] looks to provide a complete tutorial and survey about recent developments, with the objective of enabling the efficient processing of deep neural networks (DNNs). Specifically, it provides a general vision of DNNs and analyzes several hardware platforms and architectures that can run them. It also summarizes the various development resources that allow researchers and professionals to get started in the field. [19] and [20] each developed Sentiment Analysis (SA) based on experiments in different Convolutional Neural Network (CNN) configurations, with [19] implemented on Hindi movie reviews collected from newspapers and online websites. The dataset was manually annotated by three native hindi speakers for model training preparation and experiments were carried out by using different numbers of convolution layers with a variable quantity and size of filters. [21] presents a similar model to [19], with neural convolution networks. However, the original model of convolution neural networks ignores sentence structure, a very important aspect of textual sentiment analysis. For this reason, [21] adds the association of parts to the convolution neural network, which allows the model to understand the sentence structure. To counteract a lack of data and actually improve the model's generalization capacity, [21] employs a generative adversarial network to obtain the common characteristics associated with emotions. Also, [22] suggests a sentiment classification model with a convolutional neural network that uses representations of several words to represent words that have not been pre-trained. The experimental outcomes of three data sets show that the suggested model, with an additional character-level integration method, improves the accuracy of the sentiment classification. On the other hand, [23] suggests a multiple attention network (MAN) for sentiment analysis that learns word and phrase level characteristics. MAN uses the vectorial representation of the input sequence as an objective in the first attention layer to locate the words that contribute to the sentiment of the sentence.

As has been shown, there are many models and configurations of neural networks, some more effective than others depending on their application and desired analysis. This is the case with [21], as opposed to [19], because it can overcome a lack of data availability. In the case of [23], where even though an individual word can indicate subjectivity, it can give insufficient context to determine the orientation of the sentiment. The authors posit that this sentiment analysis usually requires multiple steps of reasoning. Therefore, they applied a second attention layer to explore the information of the phrase around the key word.

3 Proposed model

3.1 Model analysis

To support SMEs, we present the following framework (Fig. 1) that analyzes information from social networks on a cloud platform to create a tourism preference metric that helps create desirable tourism packages. This is very important for SME tourist agencies because their business model relies on presenting desirable travel packages for travelers to visit tourist destinations. Historically, agencies used word of mouth information and in some cases, surveys, to design their travel packages, both of which lacked reliability. With the suggested framework, information is collected from social networks and review sites and processed in a low-cost, high performance, cloud platform that uses neural networks to analyze and calculate traveler sentiment. These outcomes are stored in a datastore where metrics/reports can be generated in the future. Performance indicators and representations of place and time trends will show where tourists feel more satisfied or motivated. With this information, tourist agencies can create better packages that reflect historic travelers' mood, and a more personalized sales environment.

3.2 Components

Fig. 2 shows the process of extraction of the information from Google and Twitter, first by a filter (hashtag), then the sentiment analysis to generate the charts

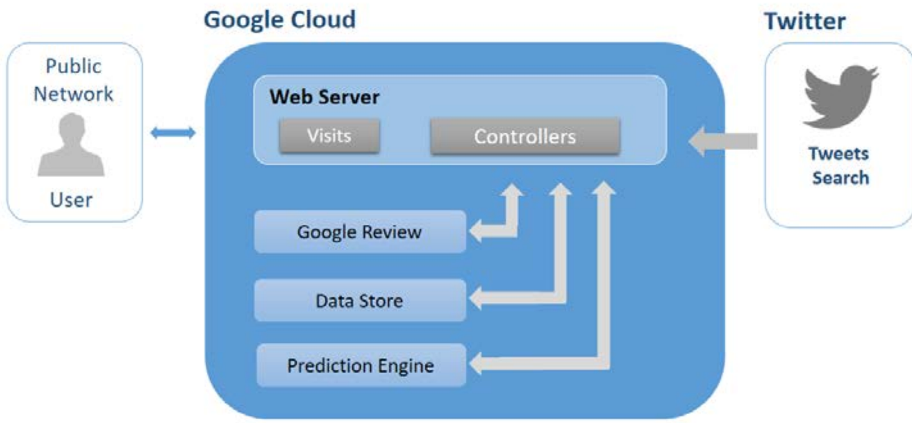

Fig. 1. Framework.

that will present the trends and tourist satisfaction in relation to the different destinations.

Input For data input we use two platforms: Twitter and Google Places/Maps. The Twitter platform will be used for measuring user motivation at a specific time, because this network fits with this type of analysis.

Twitter / motivation The motivation data source will be Twitter, a platform commonly known for sharing people's mood or opinion at a specific moment and thus effective for measuring tourist motivation at different locations in real time. This information, related to trips, tourism, etc., will be collected in a massive way using related hashtags, which will provide the first source of data about tourist destinations during specific time periods.

Google Places/Maps satisfaction The second platform that will be used is Google Places/Maps. This network will be used for quantifying user satisfaction with registered places as this network focuses on opinions and levels of customer service. Google Places/Maps generates a more historical type of information as it doesn't operate with the real time aspect of Twitter, but provides us a more specific channel for information.

Cloud To process the information, a Cloud platform will be used to help SMEs reduce operating costs (servers purchases, trainings, implementation, maintenance, etc.) and provide an market accepted processing speed and uptime standard of service.

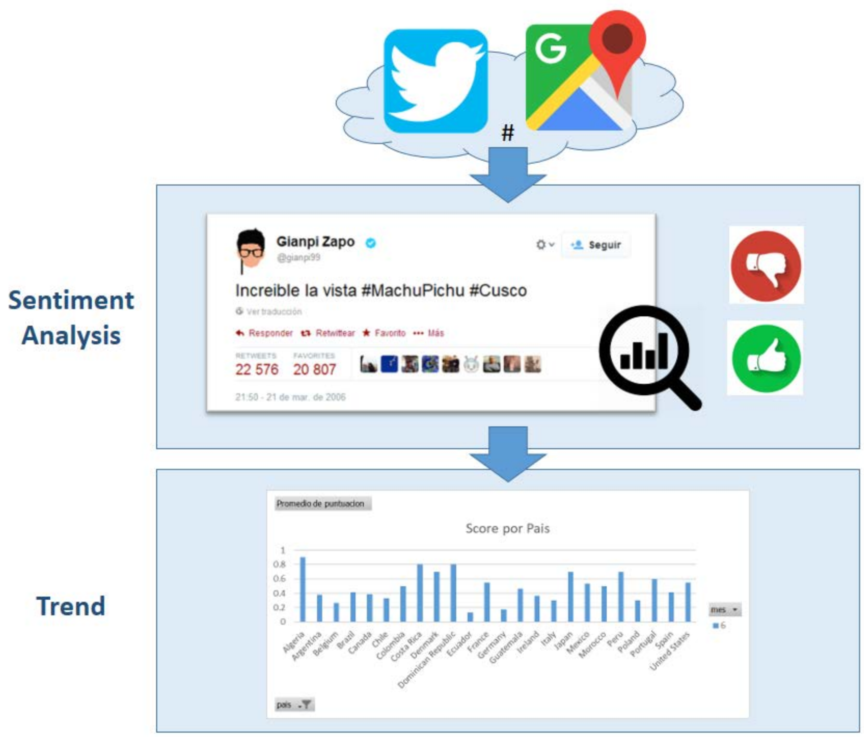

Fig. 2. Information Analysis Process.

Web server

- Tools.
  - Sent analysis. We will use neural networks for sentiment analysis on the massive information collected from Twitter and Google reviews to measure satisfaction/motivation for a specific moment and destination and ultimately quantify targeted users' moods. As the amount of data due to the number of opinions found in social networks is so large, the historical method of surveys or word of mouth opinions cannot possibly match the breadth of information from these networks. This information is vital for generating metrics and establishing trends.
  - Db. After being retrieved, the information will be stored in a database for persistence and accessibility for the application of our neural network to assess our main parameters: travel trends and popular destinations, or those destinations that present an elevated positive emotion.
  - Neural network. For the sentiment analysis we use neural networks to calculate the emotion users display on social networks. The neural network for this study is Deep Feed Forward [24]. This network combines the wide and deep models to allow a high capacity of abstraction and processing speed. This network will use the previously extracted, transformed and loaded social network input data for its analysis.
- System.
  - Analyzed data. After the social network data is organized by the neural network, it is stored in a datastore for metric and trend analysis.
  - Organized Data. At this stage the analyzed data is organized for data mining by separation/granularizating for time and location to create a decision-making spectrum for the creation of travel packages.
- Safety Support. Safety support will manage user access/communication/sessions for better organization and customer service.

Output Organized data will be presented through charts and indicators to show destination trends with respect to specific time periods to improve the selection or creation of tour packages and thus provide a better experience to customers.

Trends charts These charts will show the evolution of social network users' moods regarding the tour destinations, presenting a timeline that will help predict when the travel packages could have the best reception.

KPIs These indicators will be used to evaluate the acceptance across different destinations at specific moments in time.

4 Validation

4.1 Case

To validate the model presented, we used a case study to demonstrate that the proposal successfully solves the needs of SMEs in the Peruvian tourism market.

4.2 OT S.A.C. Company Information

OT S.A.C is a small tourist agency business that sells and distributes tour packages, with ten employees and monthly revenues of approximately $29,000. It is located in the district of Santiago de Surco, in the city of Lima, Peru. Its main suppliers are wholesale companies that provide it with a list of packages for sale and distribution. In turn, OT S.A.C. sells custom packages as requested by their clients. From its early stages, according to national regulations regarding tourist package distributors, this company benefited by its portfolio of existing clients. As new companies entered the market in the same category and with the same products, price competition began and spurred a sudden growth in the sector.

Under these circumstances, the owners and managers focused on the use of new technologies to maintain or improve sales levels. That is why we proposed this emerging technology process model to OT S.A.C.

4.3 Implementation

In Fig. 3 the ETL (Extraction, Transform and Load) process is shown. This process is used for the ex-traction of data from the social networks, the transformation of those data, and the loading of those data into the database.

4.4 Program / APIs

For the framework implementation we developed a web tool that collects data from the Twitter and Google Places/Maps social networks using their respective APIs. With the Twitter API, tweets at a specific time can be collected and filtered by hashtags chosen for their relevance/closeness to tourism topics (Fig. 4). This data is filtered and cleaned of special characters, URLs, emoticons, and other factors that can negatively affect the sentiment analysis. Then, the data is sent to the cloud service where it will be processed and stored.

All reviews will collect the geographic location information they make reference to (Fig. 5). This information will be sent afterward to the API of the sentiment analysis tool. This will evaluate each request and will send the output of the analysis to the database API for future use.

4.5 Sentiment analysis

For the implementation of sentiment analysis, 3 types of neural networks were compared: densely connected neural network (basic neural network), convolutional neural network (CNN) and short-term memory network (LSTM), which is a variant of the networks recurrent neural. To choose which type of neural network has the best performance, the 3 models were trained with the same database. The database used was [25], which contains sentences labeled with positive or negative feelings. In total there were 2748 sentences, which are labeled with a score, 1 (for positive) and 0 (for negative), as shown in (Fig. 6). The sentences come from three websites: imdb.com, amazon.com and yelp.com.

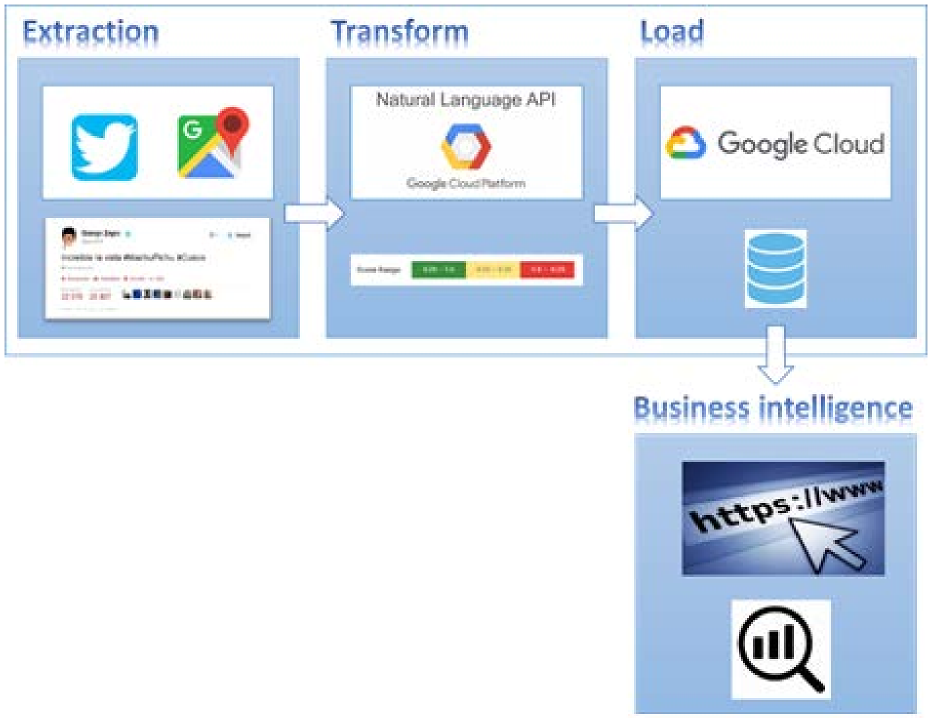

Fig. 3. ETL – Extraction, Transform, Load.

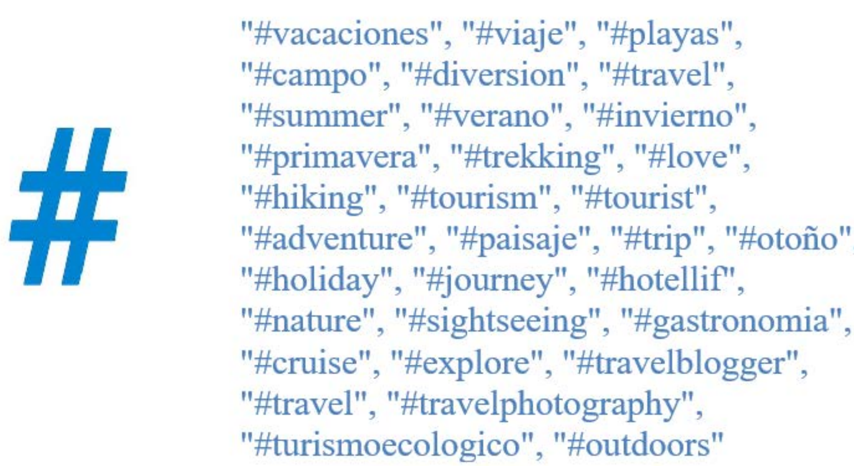

Fig. 4. Hashtags used for the filtering.

Name/ID	anio	ciudad	dia	locale	magnitud	mes	pais	puntuacion	tweet
id=4785553677484032	2019	Mercedes	16	Mercedes, Argentina	0	6	Argentina	0	#SabadoDeFlojera #Lluvia #Invierno #Casa ...
id=4787621469356032	2019	Armação dos	15	Armação dos Búzios	0.2000000029802	6	Brazil	0.100000001490	Yes, We're the couple in the airport shuttle wi...
id=4793051281096704	2019	Cuauhtémoc	10	—	2.2000000476837	6	Mexico	0.699999988079	Beautiful morning, new friends, everything is ...
id=4796213115224064	2019	Valparaiso	10	—	0.6999999880790	6	Chile	0.699999988079	Outfit #modaloretosaez #tendencias2019 #...
id=4796400818716672	2019	San Pedro Ch	10	—	1.5	6	Mexico	0.100000001490	El magnífico Popocatépetl . . . .
id=4803152373088256	2019	Chessy	14	Chessy, France	0.8000000119209	6	France	0.400000005960	Disneyland paris #travel #travelphotography ...
id=4805618086969344	2019	Caracas	10	—	0.8000000119209	6	Venezuela	0.400000005960	Habla por si sola #Lunes #Señales #AMF_C...
id=4805841895030784	2019	Boston	16	Boston, MA	1.2999999523162	6	United State	0.300000011920	Girls just want to have fun! Hanging out to ce...
id=4810099885342720	2019	La Pampa	16	La Pampa, Argentina	1.7000000476837	6	Argentina	0.5	My girl on a sunset in the countryside. @lapa...
id=4814221644660736	2019	Guadalajara	10	—	1.2999999523162	6	Mexico	0.600000023841	Mi mamá siempre me decía que antes de sal...
id=4814786122481664	2019	Almada	12	Almada, Portugal	0.8000000119209	6	Portugal	0.800000011920	Chilling 🍾 #lisboa #yosoylafigura #juniorlafi...
id=4815098690404352	2019	Almada	12	Almada, Portugal	0.8000000119209	6	Portugal	0.800000011920	Chilling 🍾 #lisboa #yosoylafigura #juniorlafi...
id=4821787967750144	2019	Tlahuiltepa	17	Tlahuiltepa, Hidalgo	0.8000000119209	6	Mexico	0.200000002980	Vi tantas Lunas... #vida #poquitaropa @DOB...

Fig. 5. Filtered, cleaned and stored tweets.

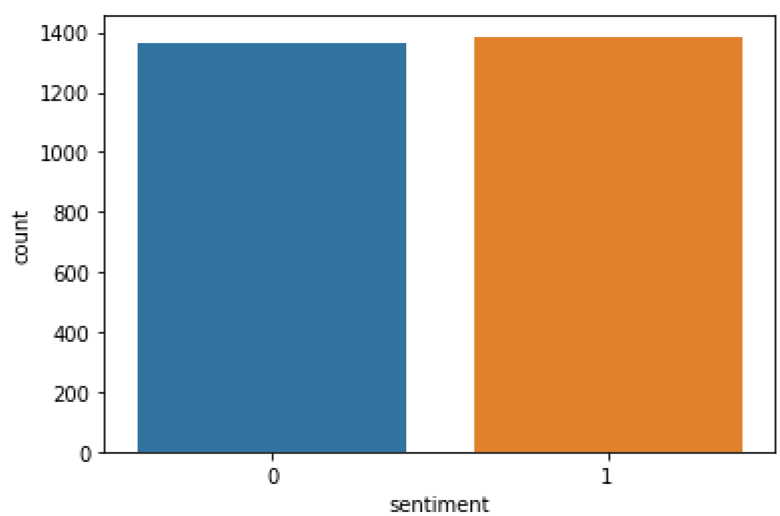

Fig. 6. Positive and negative sentences.

Densely Donnected Neural Network The first neural network to be tested is a simple deep neural network. The embedment layer will have an input length of 100, the dimension of the output vector will also be 100, and the dense layer of 10,000 parameters. For the activation function, the sigmoid function was used. The Adam optimizer was used to compile the model. For the training of the neural network, 80% of the data was used, and 20% for validation. At the end of the training, the training precision was around 81.9% and the test precision was 73.2% (Fig. 7). This means that the model is over-fitted, this occurs when the model performs better in the training set than the test set. Ideally, the performance difference between sets should be minimal.

Convolutional Neural Network (CNN) The convolutional neural network is a type of network that is mainly used for the classification of 2D data, such as images. A convolutional network tries to find specific characteristics in an image. Convolutional neural networks have also been found to work well with text data. Although text data is one-dimensional, 1D convolutional neural networks can be used to extract features from our data.

The created CNN has 1 convolutional layer and 1 grouping layer. The one-dimensional convolutional layer has 128 neurons. The kernel size is 5 and the activation function used is sigmoid. As can be seen in (Fig. 8), the training precision for CNN is 92.5%, and the test precision 83%.

Recurrent Neural Network (LSTM) Lastly, the LSTM, which is a network that works well with sequence data such as text, which is a sequence of words, will be tested. In this case, the LSTM layer will have 128 neurons, just like CNN. As can be seen in (Fig. 9), the training precision is 86% and the test precision is 85%, higher than that of CNN.

The result shows that the difference between the precision values for the training and the test sets is much smaller compared to the simple neural network and CNN. Furthermore, the difference between the loss values is also insignificant, therefore, it can be concluded that LSTM is the best algorithm for this case.

For text analysis, it is first cleaned of HTML code and symbols (Fig. 10)

Then the feeling of the phrase is predicted. the sigmoid function predicts a floating value between 0 and 1. If the value is less than 0.5, the sentiment is considered negative, and if it is greater than 0.5, the sentiment is considered positive. The result is as follows (Fig. 11):

The sentiment value for the instance is 0.87, which means that the sentiment is positive.

Finally, these results are presented in web-based statistical tables, where users can select the statistics they need for making their decisions (Fig. 12).

Fig. 13 shows a chart generated by the system, where satisfaction is shown in different countries, where 1 is a very positive comment and -1 is a very negative one. This chart has been generated from 5,000 tweets that were filtered to show only comments made in June 2019.

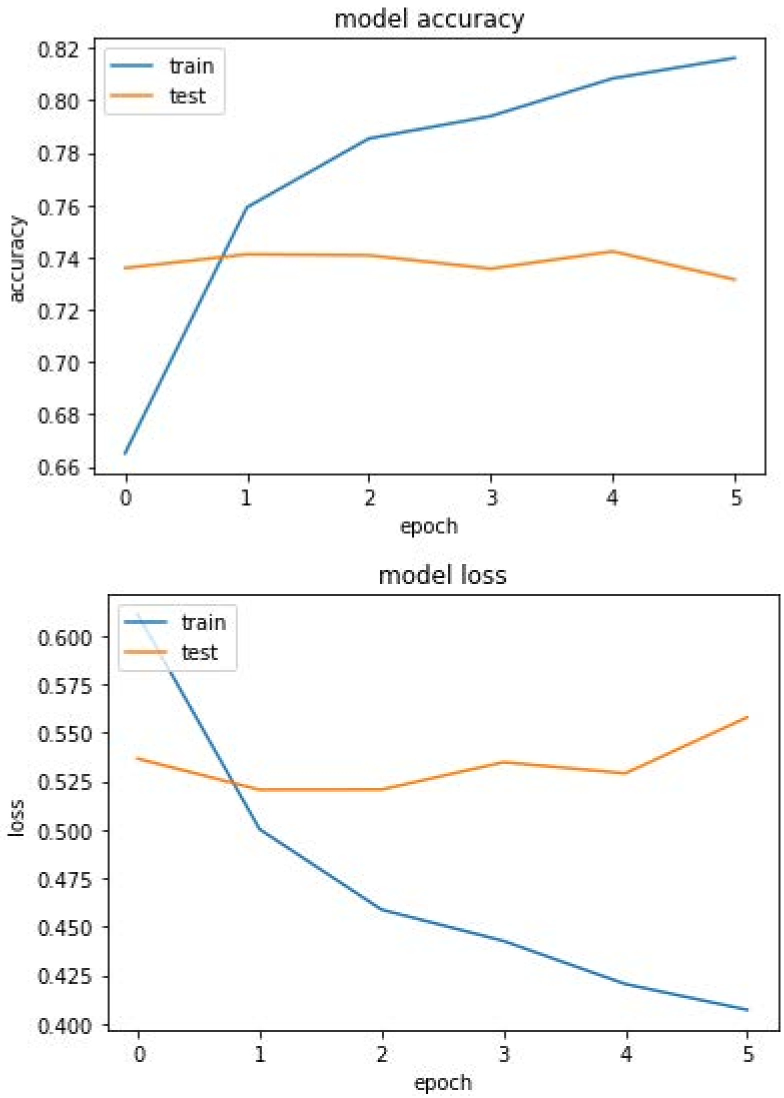

Fig. 7. Densely Donnected Neural Network.

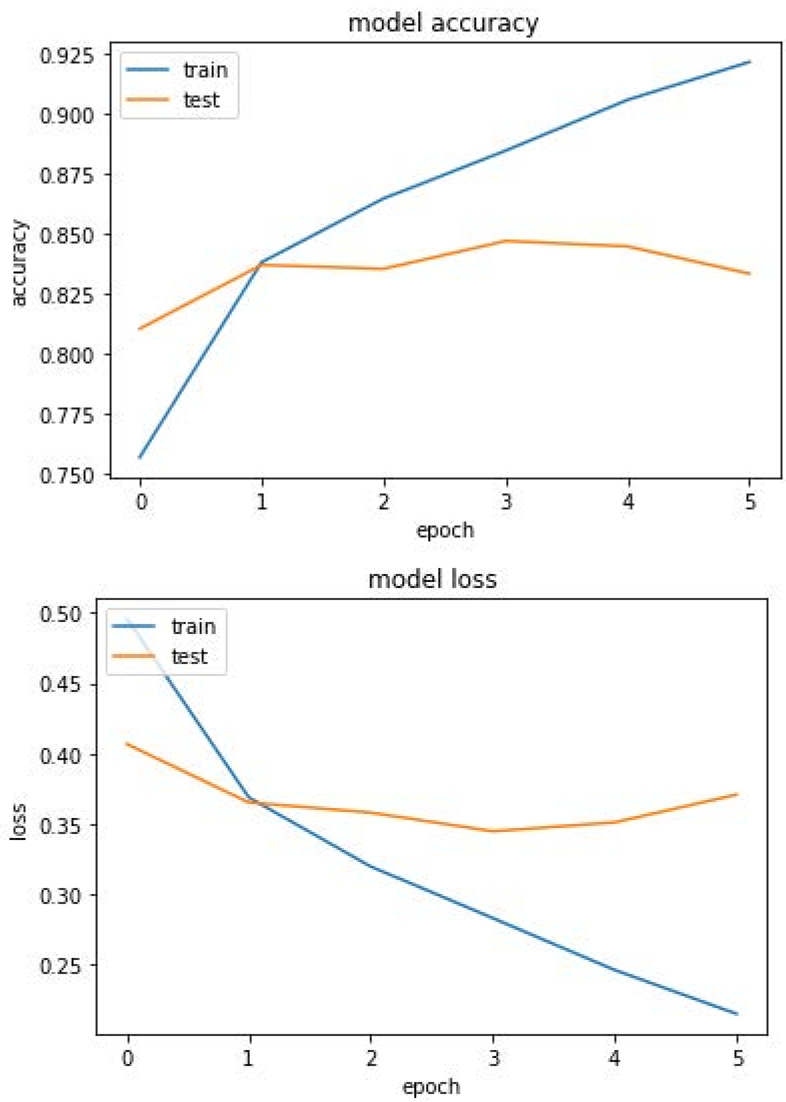

Fig. 8. Convolutional Neural Network

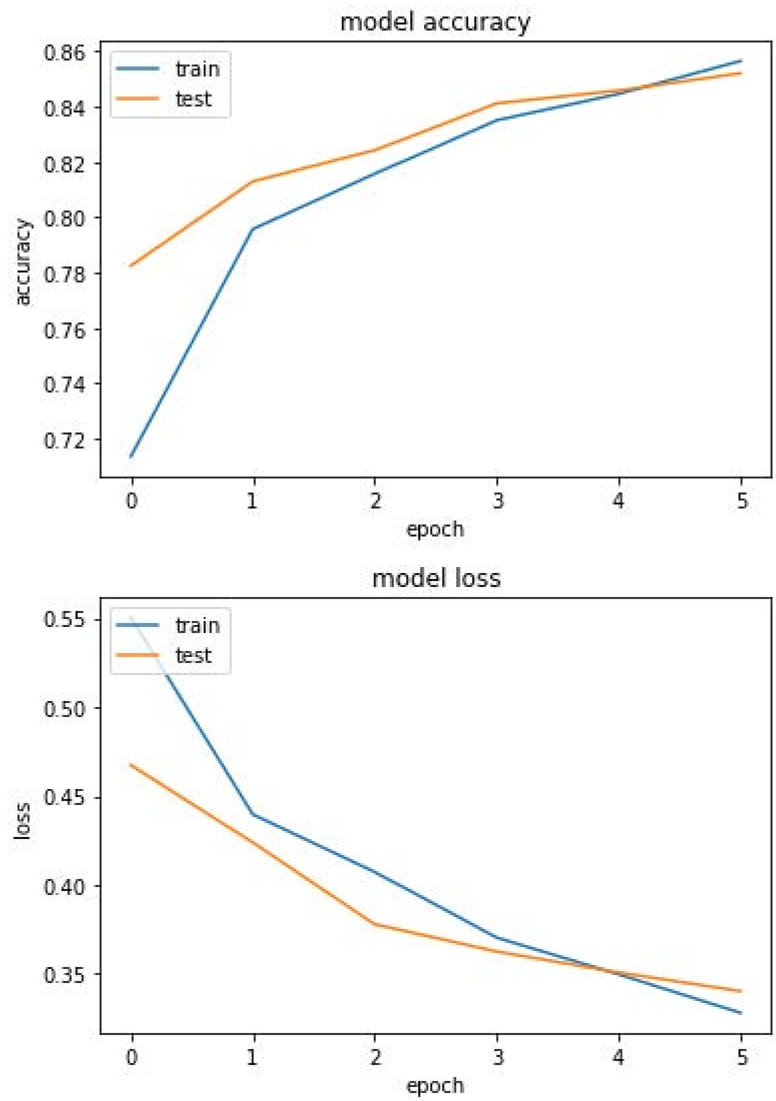

Fig. 9. Recurrent Neural Network

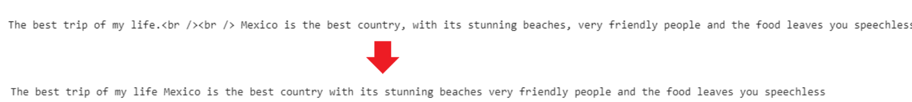

Fig. 10. Clean text.

```
array([[0.87998194]], dtype=float32)
```

Fig. 11. Result of sentiment analysis.

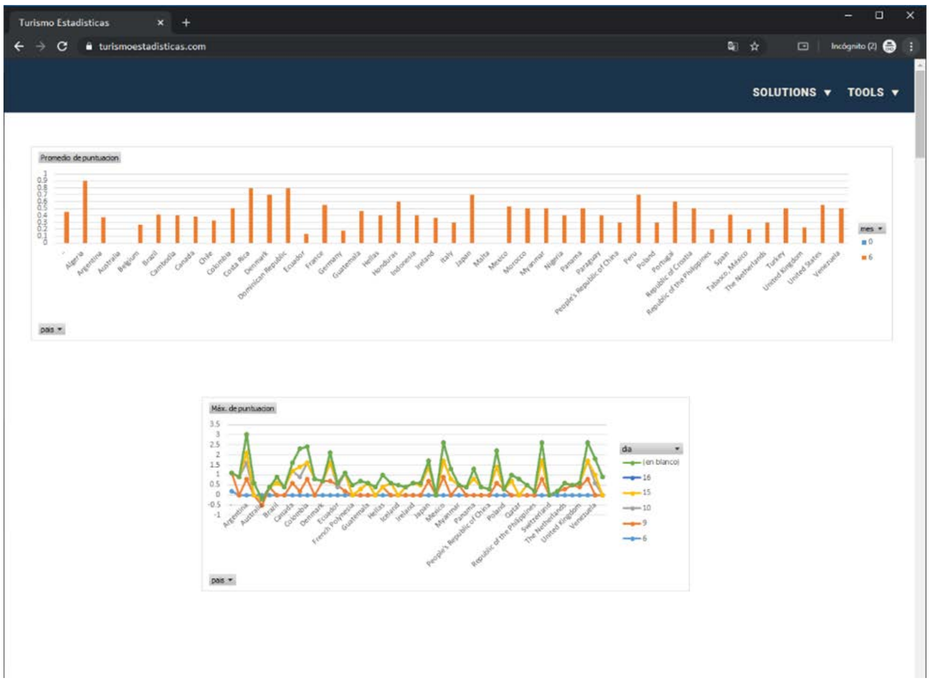

Fig. 12. Web Platform.

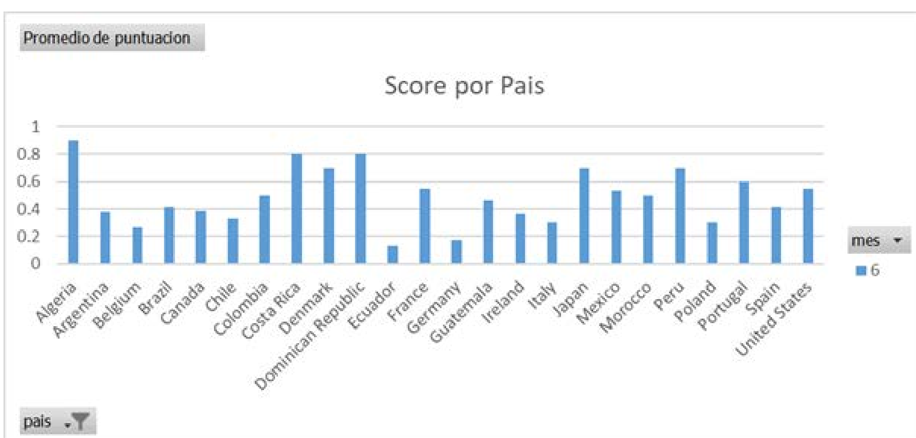

Fig. 13. Generated chart.

4.6 Implementation Outcomes

As a result of the use of the framework, data collected from social networks is processed by the sentiment analysis tool and organized so that users can understand the tourism trends that occur during specific time periods. Through charts or a trend line, users will be able to see when the highest positive sentiment occurs for a specific destination, and thus present travel packages with a high probability of desirability and acceptance. Through this framework, the travel packages presented will have a higher probability of customer satisfaction. The outcomes of Twitter analysis we will be able to rate the users' mood or motivation across time. The outcomes from Google Reviews analysis will concern satisfaction levels, helping quantify the quality level of services offered at different destinations. These two metrics, generated through the experiences shared on social network, will help agencies design better travel packages.

4.7 Segmentation

The scenarios presented were generated from the data collected throughout the project. Our first scenario presents the picture of travel and tourism sentiment from different countries. The next scenario shows the number of tweets about specific tourism topics, which reveals to us the mood of a country regarding those topics. The last shows us raw data that companies can use to build their charts and metrics. With these scenarios, we can have a wider dataset from which to analyze tourism topics from different countries across time.

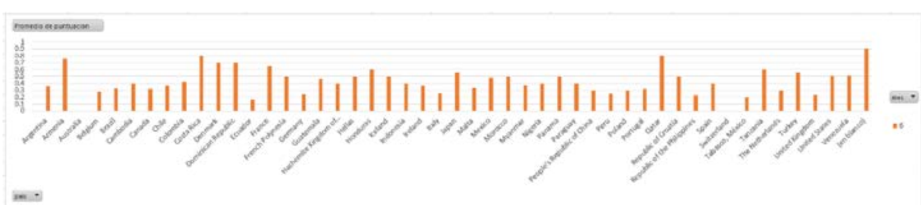

Fig. 14. Average ranking of countries per month.

Scenario 1 – Average ranking of countries per month The preceding chart shows the average ranking obtained from tweets collected in June and processed through the sentiment analysis tool. It shows what countries have presented a positive attitude related to travel and tourism on social networks. From this, we can visualize which countries have a better reception to tourism topics and can evaluate which periods of time would be better for trip planning and travel package presentation.

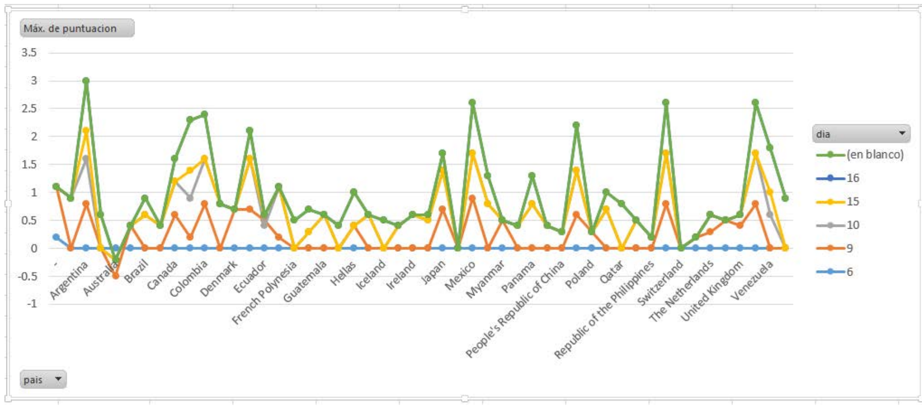

Fig. 15. Sum of the score of countries per day.

Scenario 2 – Sum of the score of countries per day In this chart, we show the sum of the sentiment analysis scores of tweets on specific days, by country. The chart is segmented by days, to present a more detailed vision of the sentiment in different countries on travel and tourism topics. The score is added to give a dimension to the number of tweets generated and to see what amount of people are discussing the topic in social networks. By taking into account the number of tweets generated and the scores obtained, we can have an idea of the tourism picture in different countries. With this, companies can make better decisions as to what are the best times to organize tourist packages.

año	ciudad	dia	magnitud	mes	pais	puntuacion	tweet
2019	A Coruña	18	0.200000003	6	Spain	0.100000001	Winter is coming!! #verano #Coruña
2019	Abbotsford	10	0.5	6	Canada	0.200000003	The Guardian Raises the Level of #ClimateAlarmism to New Heights #IPCC #WMO #climatechange #globalwarming #science #gl
2019	Acapulco de	9	1	6	Mexico	-0.200000003	Calma . . . #travelstories #traveler #traveltheworld #laguna #lagoon #acapulco #guerrero #samsunga9 #vacaciones #vacation #nub
2019	Acapulco de	16	1.600000024	6	Mexico	0.5	Buenos días acapulquito!! . . . #travel #viajar #trip #viaje #viajes #instatravel #viajeros #turismo #travelphotography #viajando #beac
2019	Acapulco de	17	0.899999976	6	Mexico	0.899999976	Los "Nachitos" OOOO #Amigos #Familia #Compayes #Fiesta #Celebración #Viaje #Cumpleaños #Junio2019 #México #Acapul
2019	Adelaide	10	0.899999976	6	Australia	0.200000003	Be curious..... about Nepal. #nepal #kathmandu #pokhara #trek #annapurnacircuit #everestbasecamp #thedontforgettravelgroup #
2019	Advanced D	10	2.799999952	6	United Stat	0.5	Are you on Summer Vacay yet? We want to hear about your #summer plans! We're celebrating Father's Day this Sunday, our 5–
2019	Alcobendas	16	0.899999976	6	Spain	0.400000006	Si vives correctamente los Sueños vendrán a ti □ #primavera #stronger #actitudpositiva #smile #smiling #Madrid #love #blogger #
2019	Alegria	17	1.399999976	6	Republic of	0.300000012	Thank God for these creation. Thank God for the experience. Thank God we all survived. #Summer #Vacay #Cebu #Can

Fig. 16. Analyzed Tweets.

Scenario 3 – Analyzed Tweets This chart shows an abstract from the obtained tweets. The table shows the time where the tweet was collected and its sentiment analysis. The Magnitude is the numeric equivalent of the intensity of the sentiment shown in the tweet, while the Score shows how positive or negative the tweet has been rated. With this information we can generate charts and metrics to help with company decision-making.

4.8 Case study result

After the process model implementation and thanks to the support of the web tool, OT SAC reduced decision-making time by 60%, to an average of 6 days (in a 3-8 day range).

Due to the support provided by the sentiment analysis, OT SAC was able to reduce logistic expenditures, as seen in Table 1. In addition to increasing travel package sales, they were able to create their own packages, reducing dependence on a wholesale company (See Table 2).

Table 1. Loss by logistic expenditures.

Loss caused by logistical expenses ($)			
Before		**After**	
Annual expenses caused by price difference	4824.24	Annual expenses caused by price difference	3600
Annual loss caused by purchases	11,104.3	Annual loss caused by purchases	7773.01
Loss %	57%	Loss %	27%
Indicator	Critic	**Indicator**	Positive

Table 2. Before and after sales comparison.

Sales comparison ($)			
Before		**After**	
Separation of Package with Wholesaler	10%	Separation of Package with Wholesaler	0%
Coordination with transport	5%	Coordination with transport	5%
Average sales	6	Average sales	13
Total price	$ 2810	Total price	$ 2810
Gain per package(20%-10%)	$ 281	Gain per package(20%)	$ 562
Total gain	**$1686**	**Total gain**	**$ 7384**

Regarding the results, we can say the following:

Before implementing the process model, the company lost customers for not responding to requirements on time, and lost package reservations due to not considering clients' needs. However, after implementation, not only could customers' needs be more effectively considered, but the company was able to negotiate more favorable rates with suppliers that will generate a greater profit margin.

5 Discussion

The neural network has been trained with sentences in English, however, it can be trained with any other language. An already analyzed database was used, since it requires a large number of already analyzed sentences to train the neural network, however, for it to perform better it is recommended to train it with more updated comments and using slangs, so that it can analyze the comments with more precision.

The proposed system does not interpret emojis, and in many comments the sentiment information is contained in these, that is why, it is recommended to assign a word to each emoji and train the network with these values so that you can later interpret the emojis.

6 Conclusions

The results show that the framework will help SME tourist agencies use historical data and sentiment analysis to offer more desirable, customized travel packages.

The generation of charts and trends will vary according to the user necessity, by place, time, etc., and is immediate since all of the information required is organized and stored in a datastore in the cloud.

Some tweets do not present any sentiment by text, but rather through a picture attached to the tweet. Further research should investigate adding a system of sentiment analysis of images to the developed text-based framework to provide an even more accurate analysis.

References

1. Confederación Nacional de Instituciones Empresariales Privadas – CONFIEP (2019). Turismo en Perú (21 Noviembre del 2019). Recuperado de https://www.confiep.org.pe/noticias/economia/turismo-en-peru/
2. Zapata, G., Murga, J., Raymundo, C., Dominguez, F., Moguerza, J.M., Alvarez, J.M. (2019). Business information architecture for successful project implementation based on sentiment analysis in the tourist sector. Journal of Intelligent Information Systems, 53 (3), pp. 563-585. https://doi.org/10.1007/s10844-019-00564-x
3. Zabha, Nur & Ayop, Zakiah & Anawar, Syarulnaziah & Erman, Hamid & Zainal, Zaheera. (2019). Developing Cross-lingual Sentiment Analysis of Malay Twitter Data Using Lexicon-based Approach. International Journal of Advanced Computer Science and Applications. 10. https://doi.org/10.14569/IJACSA.2019.0100146
4. Kincl, Tomáš & Novák, Michal & Pribil, Jiri. (2019). Improving Sentiment Analysis Performance on Morphologically Rich Languages: Language and Domain Independent Approach. Computer Speech & Language. 56. 36-51. https://doi.org/10.1016/j.csl.2019.01.001
5. Fernández Gavilanes, Milagros & Juncal-Martínez, Jonathan & Méndez, Silvia & Costa-Montenegro, Enrique & Castaño, Francisco. (2018). Differentiating users by language and location estimation in sentiment analisys of informal text during major public events. Expert Systems with Applications. https://doi.org/117. 10.1016/j.eswa.2018.09.007

6. Zvarevashe, Kudakwashe & Olugbara, Oludayo. (2018). A framework for sentiment analysis with opinion mining of hotel reviews. 1-4. https://doi.org/10.1109/ICTAS.2018.8368746
7. Gunasekar, Sangeetha & Sudhakar, Sooriya. (2019). Does hotel attributes impact customer satisfaction: A sentiment analysis of online reviews. Journal of Global Scholars of Marketing Science. 29. https://doi.org/180-195. 10.1080/21639159.2019.1577155
8. Melek Meral Anitsal, Ismet Anitsal & Selin Anitsal (2019) Is your business sustainable? A sentiment analysis of air passengers of top 10 US-based airlines, Journal of Global Scholars of Marketing Science. 29. 25-41. https://doi.org/10.1080/21639159.2018.1552532
9. Alaei, Alireza & Becken, Susanne & Stantic, Bela. (2017). Sentiment Analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research. 58. 004728751774775. https://doi.org/10.1177/0047287517747753
10. Ducange, Pietro & Fazzolari, Michela & Petrocchi, Marinella & Vecchio, Massimo. (2019). An effective Decision Support System for social media listening based on cross-source sentiment analysis models. Engineering Applications of Artificial Intelligence. 78. 71-85. https://doi.org/10.1016/j.engappai.2018.10.014
11. Suresh, Hima & S, Gladston. (2019). An innovative and efficient method for Twitter sentiment analysis. International Journal of Data Mining, Modelling and Management. 11. 1. https://doi.org/10.1504/IJDMMM.2019.096543
12. Vural, A. & Cambazoglu, Berkant & KARAGOZ, Pinar. (2012). Sentiment-Focused Web Crawling. ACM Transactions on the Web. 8. 2020-2024. https://doi.org/10.1145/2396761.2398564
13. Choochart Haruechaiyasak, Alisa Kongthon, Pornpimon Palingoon and Kanokorn Trakultaweekoon (2013). S-Sense: A Sentiment Analysis Framework for Social Media Sensing. IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP). 6–13.
14. Zafra, Salud María & Martín-Valdivia, Maria & Martínez-Cámara, Eugenio & López, L.. (2017). Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter. IEEE Transactions on Affective Computing. PP. 1-1. https://doi.org/10.1109/TAFFC.2017.2693968
15. Cai, Keke & Spangler, W. & Chen, Ying & Zhang, li. (2010). Leveraging sentiment analysis for topic detection. Web Intelligence and Agent Systems. 8. 291-302. https://doi.org/10.3233/WIA-2010-0192
16. You, Quanzeng. (2016). Sentiment and Emotion Analysis for Social Multimedia: Methodologies and Applications. 1445-1449. https://doi.org/10.1145/2964284.2971475
17. Schmidhuber, Juergen. (2014). Deep Learning in Neural Networks: An Overview. Neural Networks. 61. https://doi.org/10.1016/j.neunet.2014.09.003
18. Sze, Vivienne & Chen, Yu-Hsin & Yang, Tien-Ju & Emer, Joel. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE. 105. https://doi.org/10.1109/JPROC.2017.2761740
19. Rani, Sujata & Kumar, Parteek. (2018). Deep Learning Based Sentiment Analysis Using Convolution Neural Network. Arabian Journal for Science and Engineering. 44. https://doi.org/10.1007/s13369-018-3500-z
20. M Alharbi, A.S., Doncker, E. de (2018). Twitter Sentiment Analysis with a Deep Neural Network: An Enhanced Approach using User Behavioral Information. Cognitive Systems Research. https://doi.org/10.1016/j.cogsys.2018.10.001

21. Du, Changshun & Huang, Lei. (2019). Sentiment Analysis Method based on Piecewise Convolutional Neural Network and Generative Adversarial Network. International Journal of Computers Communications & Control. 14. 7-20. https://doi.org/10.15837/ijccc.2019.1.3374
22. Ju, Hyunjun & Yu, Hwanjo. (2018). Sentiment Classification with Convolutional Neural Network using Multiple Word Representations. 1-7. https://doi.org/10.1145/3164541.3164610
23. Du, Tingting & Huang, Yunyin & Wu, Xian & Chang, Huiyou. (2018). Multi-Attention Network for Sentiment Analysis. NLPIR 2018: Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval. 49-54. https://doi.org/10.1145/3278293.3278295
24. Zapata, G., Murga, J., Raymundo, C., Alvarez, J., Dominguez, F. (2017). Predictive model based on sentiment analysis for peruvian smes in the sustainable tourist sector. IC3K 2017 - Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 3, pp. 232-240.
25. 'From Group to Individual Labels using Deep Features', Kotzias et. al,. KDD 2015