Applying natural language processing to analyze customer satisfaction

Armin Alibasic and Tomo Popovic, *Senior Member, IEEE*

Abstract— **The aim of this paper is to analyze customer satisfaction by applying natural language processing (NLP). We have collected over 50,000 airline reviews from TripAdvisor data in the period from 2016 until 2019. This analysis demonstrates the capability of discovering the pain points of the customers by using data science techniques related to NLP. Our study shows that in today`s world, data-driven decisions must be taken quickly in order to maintain customer satisfaction and prevent customer churn.**

I. INTRODUCTION

We are living in the instant era where customers are used to getting the best service for the shortest time available. This rule could be applied generally, but it is especially applicable to the airline industry where timing, service, and hospitality are out of immense value to the customers. Continuously increasing customer expectations makes it even more challenging to maintain high customer satisfaction. Besides, maintaining good customer satisfaction could potentially lead to increasing customer profitability [1].

Today, companies receive more customer feedback than ever before. Using apps like Twitter, Facebook, Instagram, and TripAdvisor at their fingertips, customers could post a review before they have even paid for the service. As such, customer reviews provide valuable insight into various areas of the business. Hence, the brand's online reputation has the potential to either attract or deter new customers. Therefore, companies must take into consideration the received feedback and accordingly make changes in the company to improve the processes which caused complaints on the reviews.

However, the problem that arises is how to effectively and efficiently analyze thousands of reviews which are often very long texts. It will be an extremely time-consuming task if it is to be managed by human employees. Consequently, we have applied a data science technique called natural language processing (NLP) [2] to help us with this task. NLP is a mix of language, machine learning, and artificial intelligence. The goal is to learn a computer to understand human language. The methodology is described comprehensively in the following sections.

Armin Alibasic is with the University Donja Gorica, Podgorica, 81000 Montenegro (e-mail: armin.alibasic@udg.edu.me).
Tomo Popovic is with the University Donja Gorica, Podgorica, 81000 Montenegro (e-mail: tomo.popovic@udg.edu.me).

II. DATA AND METHODOLOGY

A. Data Mining

Recent vast development in computer memory, processing speed, interconnectivity, brought us to the era of Big Data. Mining the insights from user-generated data has become a hot topic among businesses due to its valuable information, e.g., ratings, complaints, suggestions, etc. However, this data is unstructured and due to its noisy nature, this type of data is the least exploited. The aim of this analysis is to get the publicly available TripAdvisor data and create knowledge from it.

For data collection, a data mining technique known as scraping is applied. This allowed us to gather publicly available TripAdvisor data from the Internet. We used the Python programming language and Beautiful Soap library [3] to gather more than 50,000 airline reviews. The main focus was on the three middle Eastern airline companies namely Etihad, Emirates, and Qatar Airways. The choice for these three was mainly due to the similar geographical area and they are competing for the same market, hence, the results will be comparable.

Then, to extract useful information from the above-mentioned data, using the R programming language, NLP tools were deployed. This field of Data Mining is referred to as Text Mining. Text mining methods are commonly used for:

- Topic Extraction - e.g. recognize the topic of web article – sport, travel, economy, etc. [4, 5],
- Sentiment Analysis - used on the web and social media monitoring to analyze the attitude and emotional state of the writer - positive, negative or neutral [6],
- Market Intelligence - track and monitor the market to extract the necessary information for businesses to build new strategies e.g. Emirates runs a marketing campaign, then Etihad gets the data immediately and can react quickly [7],
- Personalized Advertisement - placement of advertisements in the right place at the right time and for the right audience by getting insights from passengers' preferences [8].

In-detail analysis methodology has been explained in the following section B - Analysis Methodology.

B. Analysis Methodology

Below Table I shows the 50553 reviews that were collected for further analysis.

TABLE I
COLLECTED TRIPADVISOR REVIEWS

	Etihad	Emirates	Qatar
Total Reviews:	8567	28694	13292
Dates:	2016 - 2019		
Average Rating Score (1-5):	3.5	4.2	4.2

Once the data collection part was over, we started with the exploratory analysis of the data. The first step was to visualize the number of reviews per each month as shown in Fig. 1:
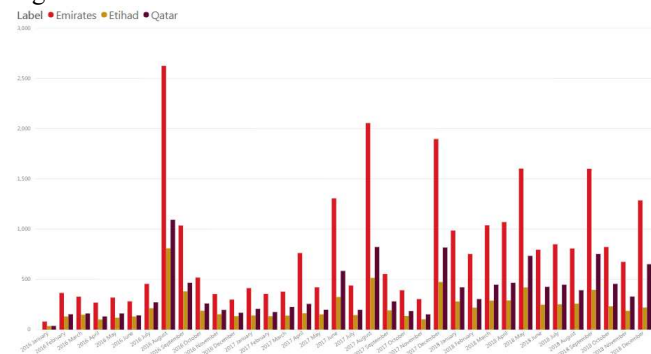

Figure 1 - Number of reviews per month

What we noticed immediately is that in certain months there are spikes in the number of reviews. For example, in August 2016 for Emirates, there was a huge spike at around 2500 reviews compared to the average of 500 reviews per month. Upon investigation, we found out that on 4th August 2016: Emirates flight EK521 was involved in an operational incident upon landing in Dubai where the plane crashed upon landing. What was expected is that most of the reviews will be negative due to the crash but to our surprise, reviews were mostly positive and people were praising Emirates on their professional reaction after the crash where all 282 passengers and 18 crew members were evacuated safely. This exploratory example showed us that the collected data is very valuable, and we continued in further data analysis.

Real-world data tend to be incomplete, noisy, and inconsistent. Hence, to ensure good data quality, before moving to the next stage of performing NLP, the following data preprocessing steps [9] were applied:
1. Tokenization – split a text into tokens (single words)
2. Removing English Stop-Words (and, or, to, the, etc.)
3. Additional Filtering (may, usually, using, often, etc.)
4. Stemming - cutting the words to the root of that word e.g. flying - fly, seats - seat, etc.

After the data processing steps were done, we were able to graph the most common words used across all the reviews. For example, Figure 2. shows the most common words in the Etihad reviews.
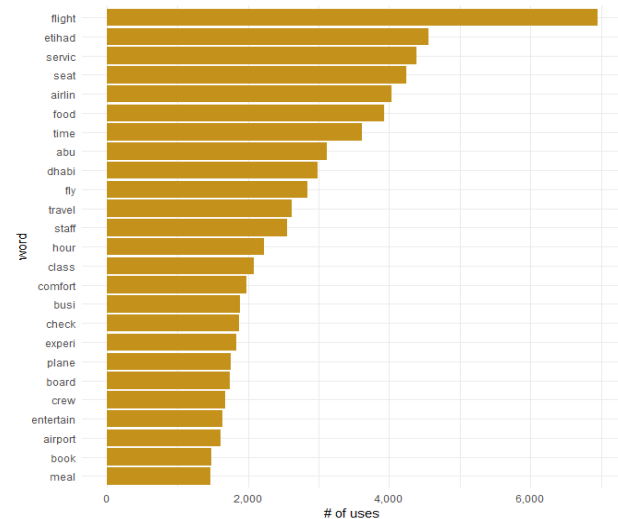

Figure 2. Most common words (Unigrams) in Etihad reviews

As expected, the highest frequency have words like flight, service (note the steamed word), seat, food, fly, staff, etc. However, one word (Unigram) by itself does not give us any valuable insight. For example, we know the people are talking about the flight or service, but the real value would be if we know what they told about that flight or service, was it good, bad, or something else. To find out that, we also explore Bigrams.

Bigrams - we often want to understand the relationship between words in a review. What sequences of words are common across review text? Figure 3. shows Bigrams for the Etihad airline.
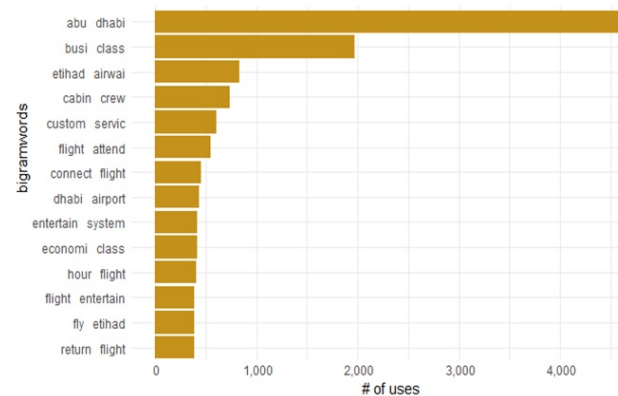

Figure 3. Etihad most common Bigrams

Again, we see that the highest frequency are words related to the airlines, and we see that people are talking about a class of travel (business or economy), cabin crew, in flight entertaining systems (IFE), food, leg space, etc. However, we are still missing the main point and that is to know are customers talking positively or negatively about these topics.

We can go even further to explore Trigrams (three most common words that come together) but as we increase the size of words that go together, the frequency of these mentions is decreasing drastically. Hence, the solution is to introduce sentiment analysis [10].

In sentiment analysis, we identified the key words from the reviews by utilizing the Python NLTK library which contains dictionary of positive and negative words. Figure 4. shows the words that contribute to the most positive and negative sentiment in the reviews.
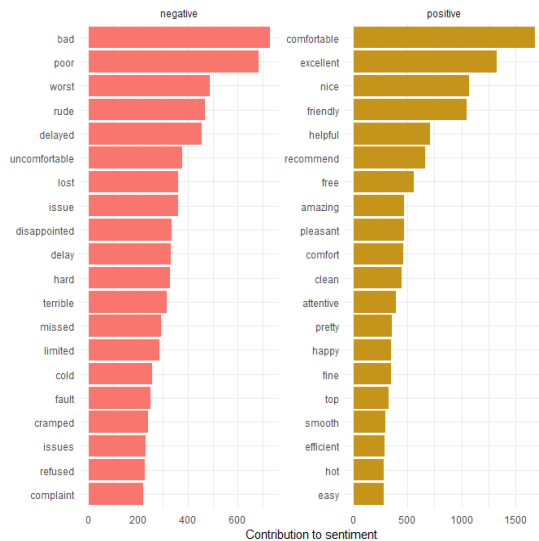

Figure 4. Words that contribute to positive and negative sentiment in the reviews

Figure 5. The most common positive or negative words to follow negations such as "no", "not", "never", "without"

Finally, we now know what people are complaining about and what they are pleased about. For example, we see that a high number of words contributing to the negative sentiment are related to rudeness, delays, lost baggage, uncomfortable seats, etc. While the compliments were for the nice, friendly and helpful staff, for clean airplanes, smooth flights, etc.

Nevertheless, analyzing single words for the sentiment analysis can be misleading. What if for example customer said "Not good" but we split words and the word "good" counts as a positive statement. On the contrary, customers can say "Not bad" and we will count bad as negative while this was a neutral review.

To quantify the misclassifications, we applied sentiment analysis using bigrams on the most common negation words such as not, never, no, without. Figure 5. shows these misclassifications in the sentiment analysis. As it can be seen, the highest number of misclassified bigrams were for the "not" word with the frequency going up to 300 examples in our dataset.

While this can be seen as a weakness to slightly mislead the overall sentiment analysis score, Figure 5. discovers some very insightful findings such as comments "no smile", "not care", "not friendly". This shows to the airline crew that such a simple thing as smiling to the customer is very important to them and can change their mood from negative to positive.

Also, as we can see we have some misclassification on the opposite side so for example a bit more than 100 times word "bad" was counted as negative, while the context was "not bad" meaning that it was a neutral or positive sentiment. Overall, in most misclassified cases, the positive and negative will tend to annul so the final results of this analysis are still valid and trustworthy.
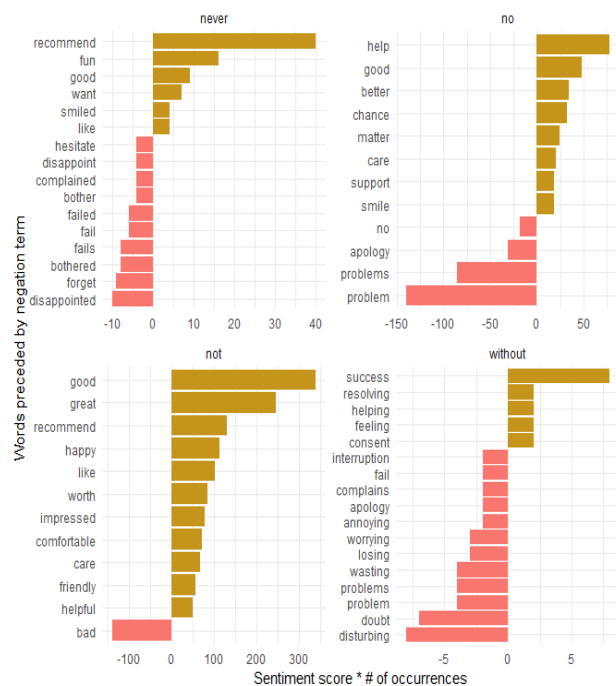
III. RESULTS

A. Sentiment Analysis

The sentiment score is calculated using the frequency of appearing positive and negative words which are matching the predefined dictionary of positive and negative words where each word will have its own score of positive (max score 5) or negative (max score -5) impact.

The score is automatically calculated for all reviews but due to space constraints, below we show sentiment scores for three reviews that represents most positive, neutral and most negative reviews.

The most positive review **id 1541 (sentiment score 3.6)**: "**Fantastic** service and food, perhaps could have done with more meals rather than one per flight. The A380 is just **superb**, also went on the 777 for one leg which I felt it was good but nowhere near to the flagship carrier. The A380 bed is **brilliant**. 777 bed so-so. Lounge in Abu Dhabi and London were **lovely** also."

Almost neutral (combination of positive and negative words) review **id 5198 (sentiment score -0.2)**: "Airplane was **small** and **old** like domestic flight. We entered boarding **late** and sit in airplane for one-hour **delay** without moving. Food is **ok** but not **amazing**. The **best** thing is they offer a shuttle bus from Abu Dhabi to Dubai and vice versa for free."

Most negative review **id 6082 (sentiment score -2.8):** "I had such a **great** flight going into Delhi in January so coming back to Toronto was **shocked** how **awful** the service was First service of meal was a brown bag with **cold** sandwich and cookies. Second service was even **worse**. My seat number was 29H for which I paid an extra $235 for more leg room."

B. Customer Ratings

Lastly, we analyzed the overall ratings for Etihad in the first five months of 2019. Then, we compared this to the same period in 2018. We found that there was a year over year (YoY) drop of 16% in the overall rating. We examined further by diving the ratings into eight different categories as shown in the Figure 6.
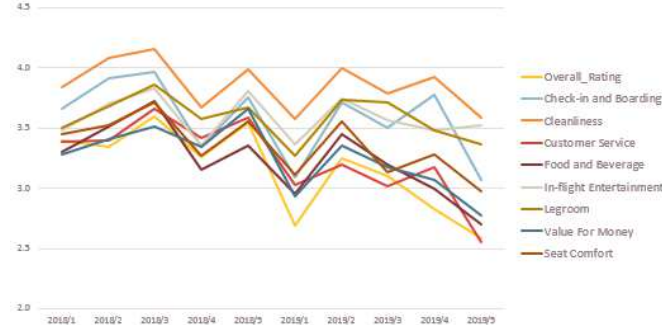

Figure 6. Etihad rating in the first five months of 2018-2019

The results showed that the category Customer Service dropped by 14% YoY, followed by the Value for Money 11% drop, and Food and Beverage 10% drop. The least drop was in the IFE of only 3% in 2019 compared to 2018.

To understand these changes even better, we have developed a PowerBI dashboard where users can easily change filters and analyze what they need as shown in Figure 7.

IV. CONCLUSION

This paper demonstrates the power of NLP and its usefulness in analyzing a large number of human generated text to drive valuable insights. It shows that from noisy unstructured data it is possible to derive conclusions and make a data-driven decision that can improve the business by improving customer satisfaction.

REFERENCES

[1] Gurău, C. and Ranchhod, A., 2002. Measuring customer satisfaction: a platform for calculating, predicting and increasing customer profitability. Journal of Targeting, Measurement and Analysis for Marketing, 10(3), pp.203-219.

[2] Bird, S., Klein, E. and Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

[3] Zheng, C., He, G. and Peng, Z., 2015. A Study of Web Information Extraction Technology Based on Beautiful Soup. JCP, 10(6), pp.381-387.

[4] Bougiatiotis, K. and Giannakopoulos, T., 2016, May. Content representation and similarity of movies based on topic extraction from subtitles. In Proceedings of the 9th Hellenic Conference on Artificial Intelligence (pp. 1-7).

[5] Alibasic, A., Simsekler, M.C.E., Kurfess, T., Woon, W.L. and Omar, M.A., 2020. Utilizing data science techniques to analyze skill and demand changes in healthcare occupations: case study on USA and UAE healthcare sector. Soft Computing, 24(7), pp.4959-4976.

[6] Yi, J., Nasukawa, T., Bunescu, R. and Niblack, W., 2003, November. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In Third IEEE international conference on data mining (pp. 427-434). IEEE.

[7] Laskowski, M. and Kim, H.M., 2016, July. Rapid prototyping of a text mining application for cryptocurrency market intelligence. In 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI) (pp. 448-453). IEEE.

[8] Lin, K.P., Lai, C.Y., Chen, P.C. and Hwang, S.Y., 2015, October. Personalized hotel recommendation using text mining and mobile browsing tracking. In 2015 IEEE International Conference on Systems, Man, and Cybernetics (pp. 191-196). IEEE.

[9] Kannan, S. and Gurusamy, V., 2014. Preprocessing techniques for text mining. International Journal of Computer Science & Communication Networks, 5(1), pp.7-16.

[10] Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp.82-89.
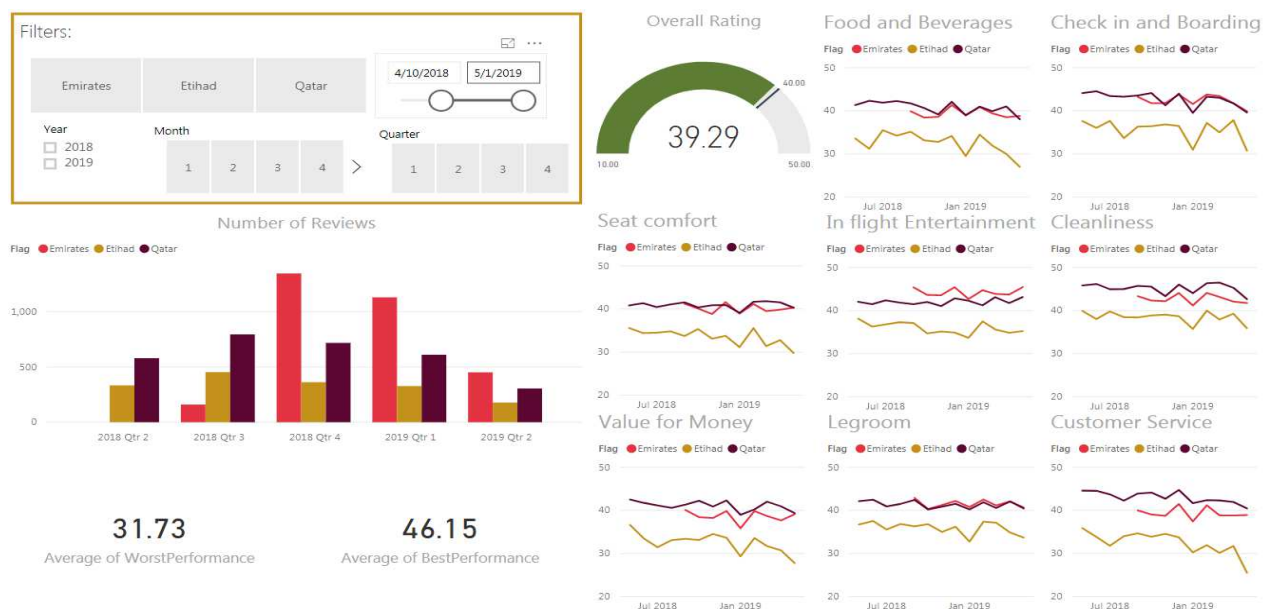
Figure 7. PowerBI dashboard for analyzing all airlines