Towards a Machine Learning and Datamining approach to identify customer satisfaction factors on Airbnb

Mohamed Chiny
Laboratory of Computer Sciences
Ibn Tofail University
Kenitra, Morocco
chiny@uca.ma

Omar Bencharef
Department of Computer Sciences
Cadi Ayyad University
Marrakesh, Morocco
o.bencharef@uca.ma

Younes Chihab
Laboratory of Computer Sciences
Ibn Tofail University
Kenitra, Morocco
younes.chihab@uit.ac.ma

2021 7th International Conference on Optimization and Applications (ICOA) | 978-1-6654-4103-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICOA51614.2021.9442657

***Abstract*— Most companies make customer satisfaction a top priority. Indeed, satisfaction is considered as one of the pillars of customer loyalty in order to bring them to use a product or a service again and the tourism sector is not an exception. In fact, all the companies working in this sector strive to provide the best possible service to their customers in order to gain their satisfaction, in this case the companies specialized in P2P accommodation rental such as Airbnb. In the existing literature there are many works that have tried to identify the identifiers of customer satisfaction. However, we have not found any study that tries to capture these identifiers by taking into account the category to which the customer belongs, namely individuals, couples and families. In our study we analyzed 100,000 reviews left by customers on the Airbnb platform towards accommodations located in London. We adopted an approach based on Natural Language Processing (NLP) to segment the reviews according to the categories of the customers, then we trained two regression models; Multiple Linear Regression and Gradient Boosting Regression to determine the weights of the 6 elementary scores noted on Airbnb. The results show that the importance given to each elemental indicator changes depending on the category the customer belongs to, and sometimes even remarkably, which can have important implications for the P2P hosting industry by refining offers according to the customer's category in order to ensure the best possible experience.**

Keywords—P2P Accommodation, Airbnb, Machine learning, Datamining, NLP

I. Introduction

With more than 4 million listings published in 2017, Airbnb which is an online platform, is the global leader in booking and renting tourist accommodation between individuals [1]. Considering the popularity that Airbnb enjoys, in addition to the nature of the P2P accommodation rental industry, customers are predisposed to believe that the offers it provides are cheaper. This is due to the fact that these offers have a global visibility that endows them with great transparency, unlike the offers proposed by traditional providers such as travel agencies [2]. On the other hand, the nature of some of the goods offered on Airbnb is special and cannot be offered in the context of traditional accommodation [3].

However, in this sector where services cannot be tested before being ordered [4], and taking into account that the services offered in this sector are conceived as expensive by consumers, recommendations and feedbacks from customers who have already benefited from a given service are crucial and have the potential to have a great influence on the purchase decision of future prospects [5,6]. Indeed, these recommendations are part of the electronic word of mouth (eWOM) seen as a kind of informal communication whose potential can go as far as reshaping the relationship between the company and the consumer [7].

Customer satisfaction is often linked to customer loyalty. A satisfied customer is generally a loyal customer with respect to a product or service, and in the tourism industry (especially accommodation), loyalty has a positive impact on long-term financial performance [8].

However, although the existing literature is rich in terms of studies that attempt to identify the factors that influence customer satisfaction in the field of P2P accommodation, in this case on Airbnb, and consequently the scores and positive reviews that the customer is likely to publish on the platform [1], we found that there is no study that attempts to examine these factors empirically, mainly if we take into account the category of which the customer belongs (individuals, couples, families).

On Airbnb, it so happens that the customer who is invited to write a review about an accommodation he has rented can rate 6 basic indicators (accuracy, cleanliness, check-in, communication, location and value) on a scale of 10 and then give an overall score that reflects the level of his overall satisfaction. In our study, we tried to find a linear relationship between the 6 elementary scores (features) and the overall score (target). We examined data that concern accommodations located in London and that were rented through Airbnb between December 2009 and April 2020. This data was collected from Inside Airbnb [9]. After cleaning and filtering the reviews from the dataset, we retained 100,000 usable scores and reviews.

One of the major problems we faced was that the collected data did not provide any indication of the category to which the customer belongs, so we adopted a text-mining approach. Indeed, we implemented Natural Language Processing (NLP) algorithms in order to segment the retained opinions according to the 3 defined categories.

Given the large amount of data to process, we adopted an approach based on Big Data and Machine Learning. Indeed, we trained two regression algorithms; Multiple Linear Regression (MLR) and Gradient Boosting Regression (GBR) in order to calculate the respective weights to each of the 6 elementary indicators rated by the customer.

978-1-6654-4103-2/21/$31.00 ©2021 IEEE

The results found in this study will bring additional clarification on the dimensions of customer satisfaction in the P2P hosting sector (especially on Airbnb). Indeed, hosts will be able to design the best possible experience for their customers taking into account their category. On the other hand, Airbnb could consider taking into account the category of the customers when writing their reviews of an accommodation. These results will also provide additional answers to the existing literature on the dimensions of customer satisfaction with offerings that fall within the P2P accommodation sector.

II. Related Works

In his study dedicated to identifying the determinants of prices and revenues in the P2P accommodation domain (in this case on Airbnb), Sainaghi [19] identified some features that could impact the rates such as host characteristics, location, customer reviews, destination characteristics and external comparison. In [20] Voltes-Dorta et al. mentioned that previous studies on rental prices on Airbnb have focused on room characteristics, host characteristics, and location factors to determine the rental price. Their approach is to consider these same factors across property types, locations, and seasons using ordinary least squares and geographically weighted regression methods. Their results showed significant differences between the determinants of apartment and house prices and suggest that they may help providers to better price their properties against the competition. Kalehbasti et al. [21] attempted to find the best price prediction model by selecting the most relevant features from dozens of others using regression algorithms. The results of the study led to a satisfactory accuracy of the proposed model.

Many studies have attempted to identify the determinants of customer satisfaction, in this case towards P2P hosting, such as [22] where the authors applied social exchange theory to investigate how the Airbnb platform influences experience and authenticity. They highlighted the importance of the platform and its features in enhancing the attractiveness of the service, or in [23] where the authors try to determine the importance of the information disclosed about the deals listed on Airbnb towards the consumers' purchasing behavior. Their results suggest that recommendations published on the platform have a significant impact on purchase intentions. However, in the literature we did not find any approach based on the segmentation of customers into categories to better identify these determinants.

III. Methodology

The methodology we followed consisted of a massive collection of reviews and scores posted by customers on the Airbnb platform through the Inside Airbnb tool. Since the reviews can be written in any language, we cleaned and filtered the data in such a way as to retain only those written in English. Then we segmented them according to the category the customer belongs to. At the end, we trained two regression models to calculate the weights of the 6 elementary scores.

A. Data collection

The data for this study was collected from Inside Airbnb [9] which is a survey website that reports and visualizes scrapped data regarding property rentals on Airbnb. The data collected is for homes located in London with reviews published between December 2009 and April 2020. The collected data contains 1,513,966 reviews left by customers and covers thousands of accommodations listed on the platform. Among the fields collected are the customer reviews, the 6 basic scores rated on a scale of 10 and the overall score rated on a scale of 100. Other fields are provided with the data such as the neighborhood, the characteristics of the accommodation, the booking schedule, the name of the host and whether he is a superhost or not, etc.

B. Data cleaning and filtering

This preliminary step consists in retaining only the reviews that are relevant and that do not have missing or inconsistent fields. Indeed, at the time of the review entry, some fields are not mandatory, and even if they are not, users can enter arbitrary data in order to bypass the entry control. On the other hand, the scrapping operation could result in empty values or fields with inconsistent data. Since the effectiveness of the study is mainly based on the data, this data must be relevant.

After the cleaning, we proceeded to filter the rewiews in such a way that only those written in English were retained. In reality, the language of the review does not really matter in our study, because it is mainly the score values that interest us the most, but we are planning to apply sentiment analysis algorithms on the reviews in a future work, that is why we proceeded to this filtering, especially since this operation is constraining in terms of computing resources. So, we adopted an approach that consists in computing the probability of the language based on the pronunciation features using the Naive Bayse algorithm with N-gram character. It would be worth mentioning that other methods can help to detect the language such as stop-words or dictionary. However, the method we adopted is faster and more accurate [14].

At the end of this step, we retained 100,000 opinions that we considered sufficient for our study.

C. Segmentation of reviews

One of the major problems we faced was that the data collected did not provide any indication of the client category. However, our study is based on the calculation of the weights of the 6 elementary indicators noted at the base of the customer category (individuals, couples, families). Therefore, we used algorithms derived from Natural Language Processing (NLP) in order to segment the opinions according to the categories.

Using syntactic dependency parsing, which highlights the syntactic structure of a sentence, and which is represented by a dependency tree in order to detect the relationships that its elements have [15,16], we sought to identify the opinions where the subjects refer to the first person singular (Fig. 1). This operation allowed us to segment the reviews that probably concern customers who are part of the "individuals".

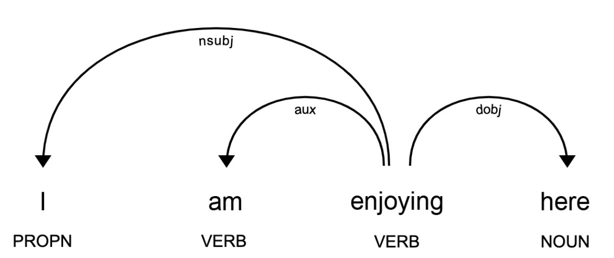

Fig. 1. The subject 'I' indicates the first person singular. This review is most likely left by a customer who belongs to the "individuals" category

This operation allowed us to segment 38,543 reviews that are most likely written by customers who belong to the "individuals" category.

As for the remaining reviews, we did not find in the literature NLP algorithms that could help us identify whether the customer belongs to the "couples" or "families" category. Therefore, we proceeded in a classical way to the processing of strings using regular expressions. In fact, we looked for the presence of occurrences that could imply whether the review is written by a customer of the "couples" or "families" category. For example, the presence of the occurrence 'my family' most probably indicates that the review is written by a customer who belongs to the 'families' category.

At the end of this operation, we identified 5,495 couple reviews and 10,874 family reviews. As for the rest of the reviews, they do not contain enough occurrences to identify their category of belonging.

We then proceeded to take a random sample of segmented reviews in order to verify their membership and thus judge the effectiveness of our approach.

D. Calculation of coefficients

In our study we used two regression algorithms (which are based on weights) to calculate the coefficients applied to each of the 6 elementary indicators rated by the customer on Airbnb.

1. Regression algorithms

Regression is a supervised learning model adapted to quantitative data. It is a method of modeling a variable (called target) according to independent predictors (called features). Regression is used when the variables are continuous (as opposed to classification models which are adapted to discrete variables). Indeed, in our case the target represented by the global score is defined in a metric space on a scale of 100, where all values between 0 and 100 are possible.

Amongst the regression algorithms we find, among others, the Multiple Linear Regression (MLR) and the Gradient Boosting Regression (GBR). Some models that were basically designed for classification can also be adapted to regression such as the Support Vector Machine for Regression (SVR), Neural Networks, Random Forest...

2. Model definition

In our study, we were interested in identifying the degree of importance of the 6 elementary indicators rated by the customer (accuracy, cleanliness, check-in, communication, location and value) with respect to the global score given by the customer and which summarizes his level of satisfaction on the whole experience. We can express the global score (GS) as a function of the elementary scores using the following formula:

$$GS = \beta_1\ AC + \beta_2\ CL + \beta_3\ CH + \beta_4\ CO + \beta_5\ LO + \beta_6\ VA$$

AC, CL, CH, CO, LO, and VA stand for accuracy, cleanliness, check-in, communication, location, and value, respectively, and β1, β2, β3, β4, β5, and β6 are the coefficients that act on the elemental scores. Our regression algorithms aim to compute these coefficients as a function of customer category. Indeed, we trained both MLR and GBR regression algorithms on the previously segmented data (38,543 reviews for individuals, 5,495 reviews for couples, and 10,874 reviews for families) by dividing them into two batches, 80% for the train set and 20% for the test set. The inputs of the model are the AC, CL, CH, CO, LO and VA features and its output is the GS target.

IV. RESULTS

After training our two models, we calculated the coefficient of determination R2, which represents a measure of the quality of the prediction of a linear regression, whose values are 0.660, 0.684, 0.712 respectively for the individuals, couples and family using the MLR model and 0.70, 0.71 and 0.74 with the GBR model.

Fig. 2 and Fig. 3 illustrate the coefficients calculated using the MLR and GBR algorithms for the three categories studied (individuals, couples and families) and which act on the 6 elementary scores noted by the client.

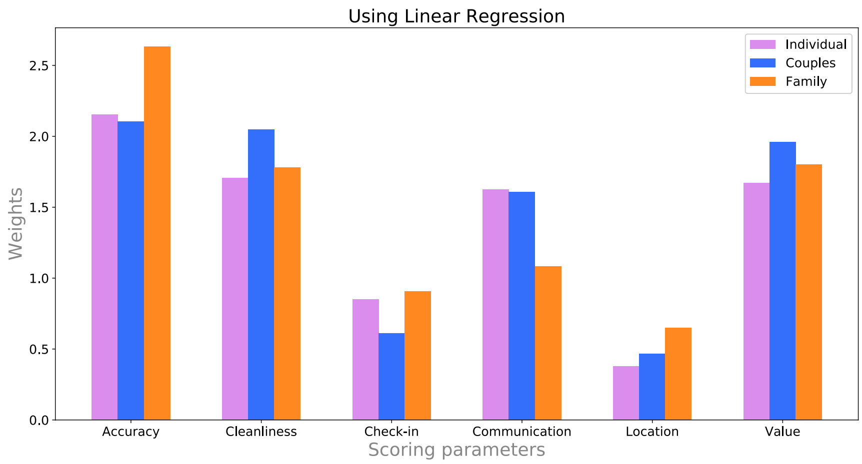

Fig. 2. Coefficients found using the MLR model

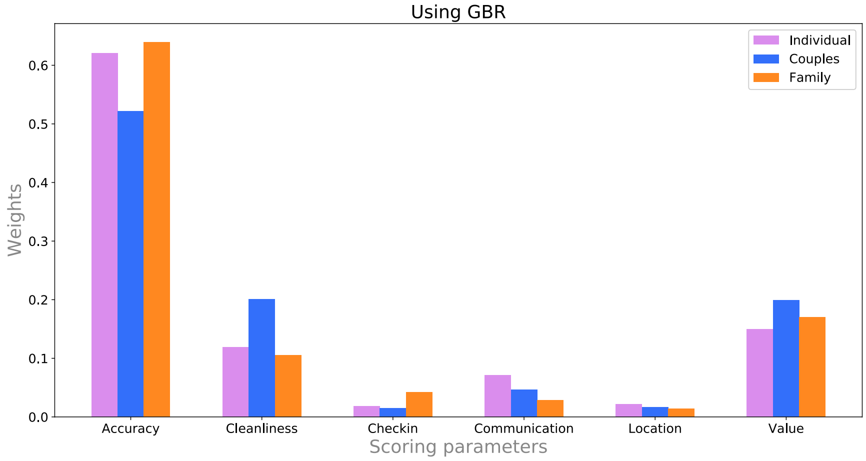

Fig. 3. Coefficients found using the GBR model

The coefficients calculated for accuracy using the GBR algorithm is clearly the largest compared to the other indicators, but overall, both algorithms give fairly consistent results with respect to the six indicators studied. The results obtained using the GBR algorithm are slightly different from those we found using the Support Vectror Machine for Regression (SVR) algorithm in another work we conducted with the same data [24].

However, after exploring these results, we can conclude that customers' expectations of P2P hosting, in this case at Airbnb, are relatively different. An overview of Figures 2 and 3 illustrates these differences, especially between the 3 customer categories studied.

Indeed, the general trend of the satisfaction indicators is almost the same for these 3 categories, even if we can notice differences between the results obtained using the MLR and GBR algorithms.

Overall, the indicators are ranked in this order: accuracy, value, cleanliness, communication, check-in and location, although there are slight differences between the results obtained with the two algorithms, if we consider each category separately. Yet, some indicators are clear, as it is the case for accuracy, which is the most important determinant for families. Indeed, Airbnb frequently recommends that hosts provide an accurate description with plenty of updated photos of the accommodations they list on the platform. An untrue description very often leads to customer dissatisfaction, and our study affirms this claim.

V. Discussion

We can state that the drivers of customer satisfaction exceed the six indicators rated on Airbnb. In another study, where we tried to propose a new centric-client evaluation system [24], we performed global score prediction using an Artificial Neural Network, and the results suggest that the 6 indicators rated on Airbnb are not sufficient to capture the dimensions of satisfaction. Indeed, the age of the customers can be a potential factor of satisfaction. Moreover, this factor has been the subject of study in many works such as [12] where the authors indicate that superior rooms and services are among the priorities of generations X and Y with a greater interest in safety experienced by generation Y. The gender [11], geographical, cultural and psychological dimensions [10,13] of customers can also have a significant impact on their evaluation.

Sometimes, when publishing online reviews of a tourist service, the customer is influenced by the functionality of the platform itself. For example, Moro et al. [17] studied the influence of the gamification features that some websites integrate, as is the case for TripAdvisor. The study found that the behavior of customers is influenced by features at the moment of writing their review, which has the effect of biasing their degree of satisfaction with their real tourism experience.

Other studies such as [11,18] have reported the influence of the amenities that the accommodation has. Indeed, most travelers want to feel at home when staying in a hotel or rented accommodation in general. For this reason, the design of the room, the Wi-Fi connection and the free food and beverages are among the factors that can positively influence the satisfaction of the customers, and thus their recommendations.

Finally, it would be absurd to claim that the overall score only depends on the 6 basic scores noted on Airbnb. In other words, these elementary scores cannot in any way summarize the indicators that influence customer satisfaction. Indeed, other indicators should be taken into account in order to accurately determine the level of satisfaction. However, when the customer evaluates a service on the platform, and the latter proposes many points to be noted, this could annoy the customer. However, the review written by the customer may contain other semantic data that could better understand his level of satisfaction by conducting a sentiment analysis, and this is precisely the direction in which our future work will be directed.

VI. Conclusion

Due to the paucity of work that attempts to elucidate the determinants of customer satisfaction in P2P accommodation, especially on Airbnb, by taking into account their category, we conducted a study where we tried to unravel these indicators according to the category the customer belongs to (individuals, couples, families). We collected a large volume of reviews left by customers who had rented accommodation via the Airbnb platform in London between December 2009 and April 2020. This data was collected from the investigative site Inside Airbnb. After cleaning and filtering the data, we proceeded to segment them according to the 3 defined categories by adopting a Natural Language Processing (NLP) based approach. Then we trained two regression algorithms; Multiple Linear Regression (MLR) and Gradient Boosting Regression (GBR) in order to compute the weights given to each of the elementary scores rated on Airbnb (accuracy, cleanliness, check-in, communication, location and value) according to the customer category.

We found that these indicators are considered by customers in the following order: Accuracy, Value, Cleanliness, Communication, Check-in, Location. We also noted that clients do not perceive these 6 indicators with the same importance depending on the category they belong to. For example, families find the accuracy of the description written by the host more important compared to the other indicators. Couples are interested in cleanliness, value, and accuracy, and individuals in communication.

However, it is very likely that the six scores noted by guests are not the only ones that influence the overall score. Indeed, dimensions that are not taken into account by this study such as age, gender, cultural and geographical dimensions of the customers, amenities offered with the accommodations, features supported by the evaluation platform... could impact the overall score attributed to the accommodation.

In the end, we believe that the results of this work could help hosts to design a better experience for their customers taking into account their category. On the other hand, Airbnb could consider adding a field that would indicate the category at the time of writing the review, so that future guests accurately apprehend the reviews in a more targeted manner. The current work also provides additional answers to the literature on the determinants of customer satisfaction especially in the area of P2P hosting.

References

[1] Liang Zhu , Mingming Cheng , IpKin Anthony Wong "Determinants of peer-to-peer rental rating scores: the case of Airbnb" International Journal of Contemporary Hospitality Management, 2019.

[2] Sérgio Moro, Paulo Rita, Joaquim Esmerado and Cristina Oliveira, "Unfolding the drivers for sentiments generated by Airbnb Experiences", International Journal of Culture, Tourism and Hospitality Research, 2019.

[3] Manojit Chattopadhyay, Subrata Kumar Mitra "Do airbnb host listing attributes influence room pricing homogenously?", International Journal of Hospitality Management, Vol 81, 54-64, 2019.

[4] Cetin, G., & Walls, A. "Understanding the Customer Experiences from the Perspective of Guests and Hotel Managers: Empirical Findings from Luxury Hotels in Istanbul, Turkey", 17th Annual Graduate Student Research Conference in Hospitality and Tourism, 2012.

[5] D. Eric Boyd, Theresa B. Clarke, Robert E. Spekman "The emergence and impact of consumer brand empowerment inonline social networks: A proposed ontology", Journal of Brand Management 21, 516–531, 2014.

[6] Eva Martin-Fuentes, Carles Mateu & Cesar Fernandez, "The more the merrier? Number of reviews versus score on TripAdvisor and booking.com", International Journal of Hospitality & Tourism Administration, 21:1, 1-14, 2018.

[7] Juan Pedro Mellinas, Sofia Reino "Average scores integration in official star rating scheme", Journal of Hospitality and Tourism Technology, 2019.

[8] Tijana Radojevic, Nemanja Stanisic, Nenad Stanic "Inside the Rating Scores: A Multilevel Analysis of the Factors Influencing Customer Satisfaction in the Hotel Industry", Cornell Hospitality Quarterly 58(2):134-164, 2017.

[9] http://insideairbnb.com/get-the-data.html, 2020.

[10] Paul Phillips, Nuno Antonio, Ana de Almeida, and Luís Nunes "The Influence of Geographic and Psychic Distance on Online Hotel Ratings", Journal of Travel Research, 2020.

[11] Vanja Bogicevic, Milos Bujisic, Cihan Cobanoglu, Andrew Hale Feinstein , "Gender and age preferences of hotel room design", International Journal of Contemporary Hospitality Management, Vol. 30(2), 2018.

[12] Fiona X. Yang, Virginia M.C. Lau "LuXurY" hotel loyalty – a comparison of Chinese Gen X and Y tourists to Macau", International journal of contemporary hospitality management, Vol. 27, 1685-1706, 2015.

[13] Sérgio Moro "Guest satisfaction in East and West: evidence from online reviews of the influence of cultural origin in two major gambling cities, Las Vegas and Macau", Tourism Recreation Research, 2020.

[14] Shuyo Nakatani "Langage Detection Librairy for Java", http://code.google.com/p/language-detection, 2010.

[15] Matthew Honnibal, Yoav Goldberg, Mark Johnson "A Non-Monotonic Arc-Eager Transition System for Dependency Parsing", Conference on Computational Natural Language Learning, 2013.

[16] Matthew Honnibal, Mark Johnson "An Improved Non-monotonic Transition System for Dependency Parsing", Conference on Empirical Methods in Natural Language Processing, 2015.

[17] Moro Sergio, Ramos Pedro, Esmerado Joaquim, Jalali Seyed Mohammad Jafar "Can we trace back hotel online reviews' characteristics using gamification features?", International Journal of Information Management, Vol. 44, 88-95, 2019.

[18] Niklas Eriksson , Asle Fagerstrøm "The Relative impact tf Wi-Fi service on young consumers' hotel booking online", Journal of Hospitality & Tourism Research, Vol. 42(7), 2018.

[19] Sainaghi Ruggero, "Determinants of price and revenue for peer-to-peer hosts. The state of the art" International Journal of Contemporary Hospitality Management, 2020.

[20] A. Voltes-Dorta, A. Sanchez-Medina "Drivers of Airbnb prices according to property/room type, season and location: A regression approach" Journal of Hospitality and Tourism, Vol. 45, 2020.

[21] P. Kalehbasti, L. Nikolenko, and H. Rezaei, "Airbnb price prediction using machine learning and sentiment analysis", 2019.

[22] TN Akarsu, P. Foroudi, TC. Melewar "What makes Airbnb likeable? Exploring the nexus between service attractiveness, country image, perceived authenticity and experience from a social exchange theory perspective within an emerging economy context" International Journal of Hospitality Management, Vol. 91, 2020

[23] Xu, X., Zeng, S., & He, Y. "The impact of information disclosure on consumer purchase behavior on sharing economy platform Airbnb" International Journal of Production Economics, 2021.

[24] Mohamed Chiny, Omar Bencharef, Moulay Youssef Hadi, Younes Chihab "A Client-Centric Evaluation System to Evaluate Guest's Satisfaction on Airbnb Using Machine Learning and NLP" Applied Computational Intelligence and Soft Computing, Vol. 2021.