



**HAL**  
open science

# Predicting A New Hotel Rating System by Analysing UGC Content from Tripadvisor: Machine Learning Application to Analyse Service Robots Influence

Jorge Calero-Sanz, Alicia Orea-Giner, Teresa Villacé-Molinero, Ana Muñoz-Mazón, Laura Fuentes-Moraleda

## ► To cite this version:

Jorge Calero-Sanz, Alicia Orea-Giner, Teresa Villacé-Molinero, Ana Muñoz-Mazón, Laura Fuentes-Moraleda. Predicting A New Hotel Rating System by Analysing UGC Content from Tripadvisor: Machine Learning Application to Analyse Service Robots Influence. *Procedia Computer Science*, 2022, 200, pp.1078-1083. 10.1016/j.procs.2022.01.307 . hal-04039212

**HAL Id: hal-04039212**

<https://hal.science/hal-04039212v1>

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



International Conference on Industry 4.0 and Smart Manufacturing

# Predicting A New Hotel Rating System by Analysing UGC Content from Tripadvisor: Machine Learning Application to Analyse Service Robots Influence

Jorge Calero-Sanz<sup>1,2,3</sup>, Alicia Orea-Giner<sup>3,4,5,6,\*</sup>, Teresa Villacé-Molinero<sup>3,5,6</sup>, Ana Muñoz-Mazón<sup>3,5,6</sup>, Laura Fuentes-Moraleda<sup>3,5,6</sup>

<sup>1</sup>Signal and Communications Theory and Telematic Systems and Computing, Rey Juan Carlos University, Madrid, Spain

<sup>2</sup>Department of Applied Mathematics and Statistics, EIAE, Technical University of Madrid, Madrid, Spain

<sup>3</sup>High Performance Research Group OPENINNOVA, Rey Juan Carlos University, Madrid, Spain

<sup>4</sup>EIREST, Université Paris 1 Panthéon-Sorbonne, Paris, France

<sup>5</sup>Business economics, Rey Juan Carlos University, Madrid, Spain

<sup>6</sup>Centro Universitario de Estudios Turísticos, Rey Juan Carlos University, Madrid, Spain

---

## Abstract

Industry 4.0 tools permit computerized creation measures, and Artificial Intelligence (AI) approaches are pivotal in investigating the travel industry. Applying these devices to decipher User Generated Content (UGC) is fundamental to understand better client's necessities, opinions, and assumptions regarding tourism services. Through this research, an exploratory analysis of results is developed through Machine Learning Models to understand better the role played by robot and traveler typologies on the rating given to hotels considering TripAdvisor reviews of 74 hotels. The purpose of this exploratory research is to develop a methodology focused on analyzing online reviews related to service robots in hotels using Machine Learning techniques to train the data collected from TripAdvisor. Preliminary results show a link between the hotel rating given in TripAdvisor and the robot typology implemented in hotels.

© 2022 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Industry 4.0 and Smart Manufacturing

*Keywords:* robot typology; traveler; TripAdvisor Rating; Machine Learning.

---

\* Corresponding author. *E-mail address:* [Alicia.orea@urjc.es](mailto:Alicia.orea@urjc.es)

1877-0509 © 2022 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Industry 4.0 and Smart Manufacturing

## 1. Introduction

Previous research on human-robot interactions (HRI) [1] confirms the importance of robots in the hotel industry and the predisposition of clients to return and recommend these hotels on platforms such as TripAdvisor. These comparative websites offer relevant information regarding the motivation of travel, the customer perception of the service provided by robots, and therefore, their satisfaction [2-3].

Advances in the use of Artificial Intelligence (AI) applied to the analysis of Big Data in the tourism sector [4] make it possible to analyse the comments of tourists on social networks applying the User Generated Content (UGC) methodology. This analysis is one of the current methods to reveal patterns, trends, and better understand user needs, feelings, and expectations of tourism services [5]. Some studies show the importance of UGC content to identify the acceptance of hotel robots [6-8]. Even Machine Learning techniques allow us to identify the impact of online reviews and predict future customer' behaviour [5]. It has been proven that both, positive and negative comments published on comparative websites, can change the customer's opinion, which means they have great influence on the purchase decision processes [8].

This research aims to make an exploratory analysis of the results derived from Machine Learning model application to analyse online reviews related to service robots in hotels collected from TripAdvisor. Following this strategy, it is feasible to anticipate the new rate given to a lodging thinking about the robot typology and type of traveller. The next research questions. The research question (RQ) is:

**RQ.** How do the different typologies of service robots and travellers can provoke and increase in the new rate given to the hotels?

## 2. Theoretical background

Analysing the customer experience is essential for understanding and improving HRI [9-10], especially in tourism services such as hotels, where the use of robots for providing services is increasingly notable [8]. This growth is due to the positive aspects of their availability [11] and the improvement of their functionalities. The necessity of studying the human-robot co-working aspects was a key element to understand the concept of Industry 5.0 [12]. Besides, an essential fact of Industry 5.0 is the analysis of HRI and the experience of final customer with the implication of robots on services [13]. As numerous studies have previously shown, tourist attitudes and perceptions are influenced by online reviews [5; 14-15], even affecting their own booking decisions and online comments [16].

In addition, customer attitudes towards robots depend on the type of robot [17] and its attribution of human characteristics [18]. According to different authors [19-20], types of robots in hotels are classified based on their functionalities as follows: (1) Front-desk robot; (2) Robot concierges; (3) Room service robots; (4) in-room robot companions; (5) cloakroom robots; (6) restaurant's robot chef and robotic bartender; (7) welcoming robots.

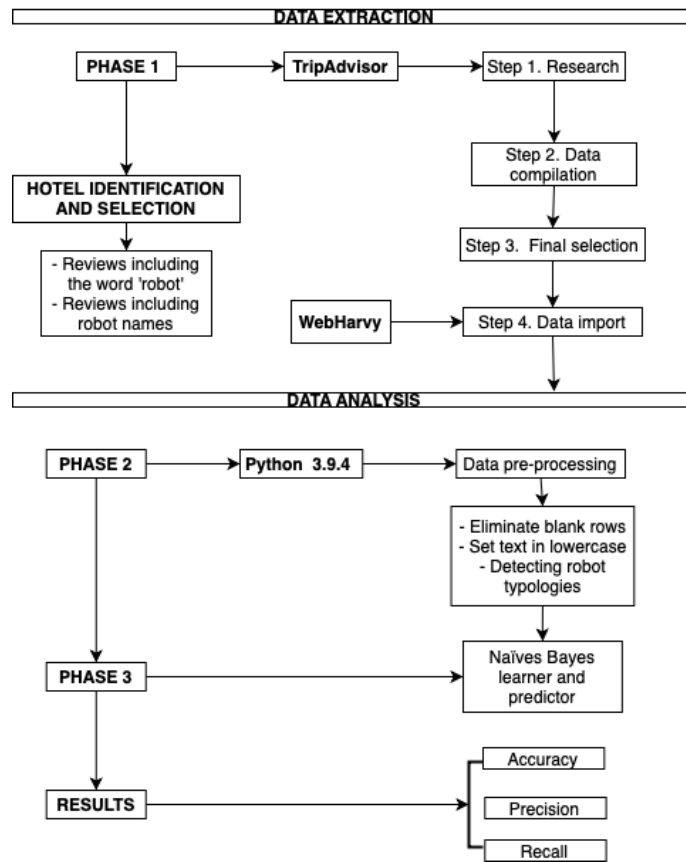
The use of Machine Learning models has been confirmed as a very useful tool to predict customer behaviour [21-22]. In the service sector, it is used to analyse how different types of robots influence on consumer perception of the Human-Robot Interaction (HRI) [23].

## 3. Methodology

This exploratory analysis gathered 8,850 online surveys including 74 lodgings utilizing robots all throughout the planet on Tripadvisor.com. These Online Travel Reviews (OTRs) [24-25] distributed in TripAdvisor have been broadly utilized by scholarly scientists in promoting, data frameworks, and inn and the travel industry writing [24-26]. This network is viewed as the world's biggest online travel analysis webpage and permits voyagers to share their whole travel insight [27]. The validity of reviews containing the experience of customers of hospitality services as data to develop a study using Machine Learning models was previously confirmed by other studies [21, 22, 27]. The site additionally offers different data, including evaluations, outlines, sorts of stay, length of visit, inn styles, lodging types, administrations, or administration appraisals (1 = poor and 5 = great).

The methodology developed presents three different phases (fig. 1). The first phase consists of the data extraction by utilizing Webharvy. Reviews including the word ‘robot’ and robot names were extracted and a preliminary dataset was created.

Fig.1. Methodology



The second phase comprises the data importation to Python and pre-processing the data to eliminated blank rows and to set the text in lowercase. After, the typologies of robots were identified by considering robot names, robot typologies and the word ‘robot’ linked to action verbs (e.g., ‘cooking robot’). The robot names distinguished were: ‘Aura’, ‘yobot’, ‘Butler’, ‘Wes’, ‘Yolanda’, ‘Wally’, ‘Emc2’, ‘Jeno’, ‘Jarvis’, ‘Leo’, ‘Pepper’, ‘Dinosaur’, ‘Alina’, ‘Winnie’, ‘Cali’, ‘Chip’, ‘Ausca’, ‘Yo2d2’, ‘Churi-Chan’, ‘Suga’, ‘Hannah’, ‘Relay’, ‘Yoshi’, ‘Botlr’, ‘Eaton’, ‘Rose’, ‘Fetch’, ‘Robi’, ‘Hazel’, ‘Botler’, ‘Dash’, ‘Bob’, ‘Holli’, ‘Hubert’, ‘Jett’, ‘Trolly’, ‘Wang’, ‘Robo bar’, ‘Chu Lee’, ‘Mr Robot’, ‘Yuri’, ‘Butlr’.. The robot names determined were identified after their usefulness dependent on the arrangement proposed by [19-20]: room service robots, restaurant’s robot chef and robotic bartender, cloakroom robots, and front-desk robot receptionist and concierge robots. Besides, the type of traveller is considered is considered as an explanatory variable. These typologies are assigned by TripAdvisor’s classification: business traveller, solo traveller, family traveller, travelling as a couple and with friends [6]. The reviews included on the final dataset were 2,230.

The final dataset is used to train machine learning models. In the third phase, the Naïves Bayes model is applied to predict the rating given to the hotel considering the robot typology. We decided to use the Naive Bayes classification method in an exploratory way due to its simplicity. The purpose is to extend the number of features

that the dataset has and to be able to train different Machine Learning models (as has been done in [5]) to analyse which one offers higher performance. By using Python 3.9.4, we applied the Scikit-Learn 0.23.1 library, applying the Naïves Bayes model to the different variables presents in our dataset. The reviews used to train the model were 70% and the 30% was used to test the model. The results obtained include the accuracy, precision, and recall.

#### 4. Preliminary Results

The Naïve Bayes modelling presents an average classification accuracy of 59.22%. This measure is linked to a higher quantity of reviews focused on a specific robot typology: 'room service robot'. It means that most of the reviews analysed containing information about the HRI relate to room service activities. Considering a previous study [6] with our results are aligned, these robots provoke a higher impact on the customer experience, and the HRI is more remarkable. For that reason, there is a preeminence of 'room service robot' typology on the dataset. The precision of the model derived from this preliminary analysis shows an average of the success rate of 37.91%. About the recall, the result is 59.22%. Finally, the F1 score is 46.2%.

The prediction shows that the values obtained are low due to the imbalance of the categories. This is because most of the reviews conducted contain a higher representation of the 'room service' robot typology.

#### 5. Conclusions, Limitations and Future Lines

Currently understanding and accurately predicting tourist behaviour is increasingly possible thanks to the use of Big Data and Artificial Intelligence (BDAI) [28]. Specifically, the use of Machine Learning methods in the hotel sector [29] allows to understand how the HRI works [30] and therefore, improve the customer experience and be more competitive. The use of robots is increasing because of the health crisis and it is an important fact to consider [31].

The findings of this study have managerial implications. First, using Machine Learning to predict future rating helps to identify how the robot typology influences customer satisfaction. Second, the predictions can identify crucial aspects of the different typologies and improve the service.

There are some limitations of this study. Firstly, the reviews obtained include the name robot to filter the comments focused on robots. We adopted this decision because it allows us to eliminate the rest of the comments about other aspects of the hotels. Secondly, Machine Learning models required large amounts of data. Even though our database includes all the reviews available about hotels and robots using our system to filter them, Machine Learning models work better with a broader range of data. However, it must be considered that most of the reviews, including HRI information, are about 'room service' robot typology. It means an imbalance of the typologies, and we must reconsider how the model was applied. This limitation can be due to significant interaction with these types of robots because customers remember their names better than other robots. Besides, it is also possible to consider that most of the robots implemented in hospitality are related to this type of functionality. Future work will be focused on reducing the typologies of robots and creating bigger groups considering their functionality.

Future research might increase our database and apply the Machine Learning models proposed and forecasting techniques to analyse the effect of the robot typology on the rating and answer the research questions presented. Besides, future research might apply more Machine Learning classifiers, as K-nearest neighbours (KNN), decision tree, logistic regression, neural network (NN), support vector machine (SVM), random forest and gradient boosting. Besides, the attributes might be increased to obtain a higher accuracy level.

#### References

- [1] de Kervenoael R, Hasan R, Schwob A, Goh E. Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots. *Tourism Management*. 2020;78:104042.
- [2] Hu N, Zhang T, Gao B, Bose I. What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*. 2019;72:417-426.

- [3] Zhao Y, Xu X, Wang M. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*. 2019;76:111-121.
- [4] Nilashi M, Bagherifard K, Rahmani M, Rafe V. A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & Industrial Engineering*. 2017;109:357-368.
- [5] Khorsand R, Rafiee M, Kayvanfar V. Insights into TripAdvisor's online reviews: The case of Tehran's hotels. *Tourism Management Perspectives*. 2020;34:100673.
- [6] Fuentes-Moraleda L, Díaz-Pérez P, Orea-Giner A, Muñoz- Mazón A, Villacé-Molinero T. Interaction between hotel service robots and humans: A hotel-specific Service Robot Acceptance Model (sRAM). *Tourism Management Perspectives*. 2020;36:100751.
- [7] Yu C. Humanlike robots as employees in the hotel industry: Thematic content analysis of online reviews. *Journal of Hospitality Marketing & Management*. 2019;29(1):22-38.
- [8] Park S. Multifaceted trust in tourism service robots. *Annals of Tourism Research*. 2020;81:102888.
- [9] Tussyadiah I, Park S. Consumer Evaluation of Hotel Service Robots. *Information and Communication Technologies in Tourism 2018*. 2017;:308-320.
- [10] Ivanov S, Gretzel U, Berezina K, Sigala M, Webster C. Progress on robotics in hospitality and tourism: a review of the literature. *Journal of Hospitality and Tourism Technology*. 2019.
- [11] Belanche D, Casaló L, Flavián C. Artificial Intelligence in FinTech: understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*. 2019;119(7):1411-1430.
- [12] Demir, K. A., Döven, G., & Sezen, B. Industry 5.0 and human-robot co-working. *Procedia computer science*. 2019;158, 688-695.
- [13] Longo, F., Padovano, A., & Umbrello, S. (2020). Value-oriented and ethical technology engineering in industry 5.0: a human-centric perspective for the design of the factory of the future. *Applied Sciences*. 2020;10(12), 4182.
- [14] Park S, Nicolau J. Asymmetric effects of online consumer reviews. *Annals of Tourism Research*. 2015;50:67-83.
- [15] Phillips P, Zigan K, Santos Silva M, Schegg R. The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*. 2015;50:130-141.
- [17] Tung V, Law R. The potential for tourism and hospitality experience research in human-robot interactions. *International Journal of Contemporary Hospitality Management*. 2017;29(10):2498-2513.
- [18] Murphy J, Gretzel U, Pesonen J. Marketing robot services in hospitality and tourism: the role of anthropomorphism. *Journal of Travel & Tourism Marketing*. 2019;36(7):784-795.
- [19] Lu L, Cai R, Gursoy D. Developing and validating a service robot integration willingness scale. *International Journal of Hospitality Management*. 2019;80:36-51.
- [20] Tussyadiah I, Zach F, Wang J. Do travelers trust intelligent service robots?. *Annals of Tourism Research*. 2020;81:102886.
- [21] Zhang, K., Chen, Y., & Li, C. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management*. 2019;75, 595-608.
- [22] Arefieva, V., Egger, R., & Yu, J. A Machine Learning approach to cluster destination image on Instagram. *Tourism Management*. 2021; 85, 104318.
- [23] Go H, Kang M, Suh S. Machine learning of robots in tourism and hospitality: interactive technology acceptance model (iTAM) – cutting edge. *Tourism Review*. 2020;75(4):625-636.
- [24] Marine-Roig E. Measuring Destination Image through Travel Reviews in Search Engines. *Sustainability*. 2017;9(8):1425.
- [23] Marine-Roig. Destination Image Analytics through Traveller-Generated Content. *Sustainability*. 2019;11(12):3392.
- [24] Xiang Z, Schwartz Z, Gerdes J, Uysal M. What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*. 2015;44:120-130.
- [25] Schuckert M, Liu X, Law R. Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*. 2015;32(5):608-621.
- [26] Lee J, Benjamin S, Childs M. Unpacking the Emotions behind TripAdvisor Travel Reviews: The Case Study of Gatlinburg, Tennessee. *International Journal of Hospitality & Tourism Administration*. 2020;:1-18.

- [27] Valdivia, A., Luzón, M. V., & Herrera, F. Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*. 2017; 32(4), 72-77.
- [28] Samara D, Magnisalis I, Peristeras V. Artificial intelligence and big data in tourism: a systematic literature review. *Journal of Hospitality and Tourism Technology*. 2020;11(2):343-367.
- [29] Claveria O, Monte E, Torra S. Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*. 2015;:1-4.
- [30] Leung X. Technology-enabled service evolution in tourism: a perspective article. *Tourism Review*. 2019;75(1):279-282.
- [31] Shen, Y., Guo, D., Long, F., Mateos, L. A., Ding, H., Xiu, Z., ... & Tan, H. Robots under COVID-19 pandemic: A comprehensive survey. *Ieee Access*, 2020.