A NAIVE-BAYES STRATEGY FOR CLASSIFYING CUSTOMER SATISFACTION: A STUDY BASED ON ONLINE REVIEWS OF HOSPITALITY SERVICES

Authors:

Manuel J. Sánchez-Franco

Business Administration and Marketing, University of Sevilla.
Avda. Ramón y Cajal n1, 41018-Sevilla, Spain.
Email: majesus@us.es

Antonio Navarro-García (*)

Business Administration and Marketing, University of Sevilla.
Avda. Ramón y Cajal n1, 41018-Sevilla, Spain.
Email: majesus@us.es

Francisco Javier Rondán-Cataluña

Business Administration and Marketing, University of Sevilla.
Avda. Ramón y Cajal n1, 41018-Sevilla, Spain.
Email: majesus@us.es

(*) Corresponding author.

Acknowledgments

This study was partially supported by Junta de Andalusia, Spain, Research Group: SEJ494.

A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services

Abstract

This research assesses whether terms related to guest experience can be used to identify ways to enhance hospitality services. A study was conducted to empirically identify relevant features to classify customer satisfaction based on 47,172 reviews of 33 Las Vegas hotels registered with Yelp, a social networking site. The resulting model can help hotel managers understand guests' satisfaction. In particular, it can help managers process vast amounts of review data by using a supervised machine learning approach. The naive algorithm classifies reviews of hotels with high precision and recall and with a low computational cost. These results are more reliable and accurate than prior statistical results based on limited sample data and provide insights into how hotels can improve their services based on, for example, staff experience, professionalism, tangible and experiential factors, and gambling-based attractions.

Keywords: Online reviews, hospitality services, customer satisfaction, text analytics, classification models, naive Bayes classifier

1. Introduction

With the increased popularity of online bookings, 53% of travelers state that they would be unwilling to book a hotel that had no reviews (Prabu, 2014). Ye, Law, Gu, and Chen (2011) report that a 10% increase in travel review ratings would increase bookings by more than 5%. In restaurants, an extra half-star has a significant effect on the popularity of a service, creates satisfaction and trust in a product, and, consequently, helps restaurants sell out 19% more often (cf. Clemons, Gao, & Hitt, 2006; Duan, Gu, & Whinston, 2005; Ghose & Ipeirotis, 2011; Park, Lee, & Han, 2007; Zhang, Ye, & Law, 2011; Zhu & Zhang, 2010). Therefore, 75% of travelers worldwide consider electronic word of mouth (eWOM) an essential information-source when planning their trips (cf. Chong, Khong, Ma, McCabe, & Wang, 2018; Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004). It is quickly and easily available to anyone and remains accessible over time (Gretzel & Yoo, 2008; Litvin, Goldsmith, & Pan, 2008; Manes & Tchetchik, 2018; Reza Jalilvand & Samiei, 2012). Although online reviews are poorly structured, they are helpful in discerning guests' preferences and demands and in understanding what makes guests return or avoid a hotel. Users become fans or friends of other users and can rate or compliment other users' reviews.

The concept of eWOM is defined as "all informal communications directed at consumers through Internet-based technology related to the usage or characteristics of particular goods and services, or their sellers" (Litvin et al., 2008, p. 4). It is an opportunity to have indirect experiences, enhance market transparency by freely sharing guests' experiences, ratings, or knowledge, reduce search costs for customers, develop hotels' online reputations, and efficiently attract guests, thereby enhancing (or, conversely, undermining) long-term online relationships and genuine customer loyalty (Gretzel & Yoo, 2008; Park et al., 2007; Tsao, Hsieh, Shih, & Lin, 2015; Ye et al. 2011). Likewise, eWOM exerts a greater influence than traditional word of mouth

(WOM) (Pan & Zhang, 2011) because of its speed, convenience, one-to-many reach, and absence of face-to-face human pressure (Sun, Youn, Wu, & Kuntaraporn, 2006). Exposure to online reviews therefore enhances users' consideration of hotels based on awareness of the hotel and attitudes toward the hotel when forming opinions (Vermeulen & Seegers, 2009).

Academic research offers several non-traditional approaches to reducing the huge variety of terms employed by guests and discovering themes about hospitality services as abstract notions or distributions over a fixed vocabulary (cf. Guo, Barnes, & Jia, 2017; He et al., 2017; Ren & Hong, 2017; Sánchez-Franco, Navarro-García, & Rondán-Cataluña, 2016; Sánchez-Franco, Roldán, & Cepeda, 2018). Individuals analyze reviews by focusing not only on the summary (i.e., star ratings) but also on the content of customers' *natural* texts based on subjectively experienced intangibles (Chaves, Gomes, & Pedron, 2012; Serra-Cantallops & Salvi, 2014). In this study, we analyze an initial dataset of 47,172 reviews of hotels in Las Vegas (USA) that are registered on Yelp. Yelp is a review aggregator of travel-related contents, much like TripAdvisor or Trivago. We identify the most relevant features of the hospitality encounter to classify ratings based on guests' post-trial experiences. We apply a machine learning approach based on extracting the best features of the reviews to classify online hotel reviews as positive or negative. We do so using the naive Bayes method (cf. Duda & Hart 1974; Langley, Iba, & Thompson, 1992).

2. Theoretical framework and research questions

Infomediation systems generate a significant amount of free online content on subjectively experienced intangible goods or experiences. They also transform how individuals search for interesting or unique experiential services and modify how travelers select hotel accommodation. Therefore, infomediation systems are a critical success factor to compete in the hospitality industry (Raguseo, Neirotti, & Paolucci, 2017; Sparks, So, & Bradley, 2016). Because of the high costs typically associated with investment in the hospitality industry, it is sensible to study the service features that travelers describe, relive, reconstruct, and share in their online reviews (cf. Sánchez-Franco et al., 2018).

Online reviews are considered a key source of relevant user-generated content (UGC) (cf. Pan, MacLaurin, & Crotts, 2007; Racherla, Connolly, & Christodoulidou, 2013). This UGC plays an increasingly important role in consumer attitudes and purchase intentions, particularly in relation to travel services (Litvin et al., 2008; Liu & Park, 2015; Marchiori & Cantoni, 2015; Wu, Shen, Fan, & Mattila, 2017). On Yelp, reviewer satisfaction is represented by the number of stars awarded by the reviewer. These star ratings can be interpreted as an overall assessment of the customers' post-consumption experience. Such ratings are so critical that an extra half-star allows restaurants to sell out more frequently. However, individuals also analyze reviews by focusing on the UGC (Sánchez-Franco et al., 2016, 2018). The UGC provided by *actual travelers* tends to be more independent, empathetic, and trustworthy (Wilson & Sherrell, 1993; cf. Bickart & Schindler, 2001; Gretzel & Yoo 2008; Huang & Chen, 2006; Sparks et al., 2016). Searching for information that is relevant to travelers' plans, including destinations, attractions, amenities, and accommodation such as hotels, guesthouses, and retreats, has thus become an indispensable step in the decision-making process.

Online reviews describe travelers' experiences of staying in hotels and reflect travelers' assessments of these hotels. These reviews are based on the way travelers live, think, and enjoy life in attractive destinations, and they reflect customers' satisfaction with the corresponding staying experience. Satisfaction is defined here as a guest's perception of how well needs, goals, and desires have been met (cf. Oliver 1999; Yoon & Uysal, 2005, for a detailed review). Satisfaction, which is related to the service provider's performance, is indeed a key measure of a hotel’s effectiveness at outperforming others. If travelers are satisfied with what they experience from hospitality services, their revisit intentions may be positively influenced (Park et al., 2007). More satisfied guests have higher-quality relationships with hospitality providers (cf. Dorsch, Swanson, & Kelley, 1998).

We adopt a product-feature-oriented approach to address two goals. The first is to explore the effectiveness of applying features of the guest experience to predict how well the whole relationship meets guests’ expectations and predictions (Sánchez-Franco et al., 2018). The second is to understand how customer feedback and complaints can be used to improve customer satisfaction. Three assumptions are made. The first is that guests’ satisfaction plays an important role in enhancing a hotel’s demand based on the number of Internet users and their frequent interactions (Xu & Li, 2016). The second is that guests’ satisfaction has a profound impact on both consumers and business organizations (Niu & Fan, 2018). The third is that guests’ satisfaction consequently leads to improved financial performance (cf. Sparks & Browning, 2011) and higher efficiency (cf. Assaf & Magnini, 2012). Based on these assumptions, the purpose of this study is to identify differences in the perceived importance of features of the guest experience and in travelers’ satisfaction with the service. The following research questions are therefore addressed: What makes a good (or bad) hotel? What are the major concerns or features (e.g., in-room facilities, cleanliness, or staff skills) and supplementary services (e.g., dining and fitness facilities) that guests consider when assessing a hospitality service? Answering these questions is essential to understand travelers’ perceptions of hospitality services and, more specifically, to design and implement systems to classify online reviews. Analyzing the text content of online reviews can provide fresh insight into what constitutes a satisfying experience.

3. Method

3.1. Data collection

We analyzed data from the 9th Yelp Dataset Challenge to understand guests' preferences, filtering records whose category contained the term "Hotels". The Yelp dataset was imbalanced in terms of location. The city of Las Vegas (Nevada, USA) was selected to remove noisy data (e.g., data containing inconsistent features based on different destinations). It was also selected because of its prestigious reputation for hosting festivals and events and its competitive advantage in entertainment driven by tourism and gambling pleasure (Douglass & Raiento, 2004; Rowley, 2015). The number of tourists visiting Las Vegas grew from 35.1 million in 2002 to about 40 million in 2016 (World Travel and Tourism Council, 2017). Las Vegas generates controversial feelings capable of affecting tourists' perceptions, as reflected in guests' opinions.

Our dataset contained 47,172 reviews of hotels for the period March 2005 to January 2017. We used the textcat package based on the R 3.4.1 statistical tool to recognize English in the reviews (cf. Hornik et al., 2013). We thereby avoided bias and kept the language variable consistent across texts. We also restricted the number of classes and considered 1-star (7,879 reviews) and 5-star (8,960 reviews) reviews (cf. Mudambi & Schuff, 2010) to sharpen the focus of our analysis. The slight imbalance in scores may owe to the guests' tendency to make positive rather than negative evaluations.

3.2. Data analysis

The Yelp dataset contained a considerable amount of plain data on businesses, users, and reviews. It was pre-processed to cleanse the data and identify trends (cf. Krippendorff, 2012; Raychaudhuri, Schütze, & Altman, 2002) by applying R 3.4.1 on macOS High Sierra with an Intel Core i5 (1.8 GHz) and 8 GB of memory. First, we discarded punctuation, capitalization, digits, and extra whitespace, tokenized and depluralized the terms, removed a list of common stop words to filter out overly common terms, and eliminated non-English characters. We also omitted terms shorter than a minimum of three characters and stemmed terms using the Porter stemming algorithm. Second, we selected a dictionary based on tf-idf values (cf. Blei & Lafferty 2009; Grün & Hornik, 2011; Salton & McGill, 1986) above the median (plus the terms "bed" and "staff"). The tf-idf approach increased proportionally to the number of times that a term appeared in a review, but it was offset by the frequency of the term in the corpus. This process yielded a vocabulary of 424 terms. Third, we built a document term matrix (DTM). A DTM is defined as a bag of words to represent text where every term is conceptualized as a feature that occurs in a document (Wallach, 2006; Zhang, Rong, & Zhou, 2010). For training and testing, the DTM was divided as follows: 80% (13,472 reviews) for training and 20% (3,367 reviews). Testing was conducted by applying the caret package (version 6.0-78; cf. Kuhn, 2008). We trained and tested the naive Bayes classifier using the e1071 R package (created by Meyer et al., version 1.6-8).

Next, an automatic text classifier was built using a learning process from the set of pre-labeled reviews to select features capable of discriminating samples that belonged to different classes (e.g., very positive 5-star ratings or very negative 1-star ratings) and identify the class to which a new review belonged. Naive Bayes is based on Bayes' theorem. It offers "a competitive classification performance for text categorization compared with other data-driven classification methods" (Forman, 2003; Genkin, Lewis, & Madigan, 2007; Tang & He, 2015; Yang & Pedersen, 1997; cf. Tang, 2016, p. 2509). Classifying non-structured text is a complex task because of the high dimensionality of the feature space and the high level of redundancy that may exist in the bulk of available data. Although the high dimensionality and redundancy cause poor predictive performance, feature (i.e., term) selection in text categorization in this study aimed at identifying the most relevant, explanatory input terms in the dataset to improve performance and increase the comprehensibility of the mining results.

We applied a filter-based approach to identify significant features and discard redundant features. A filter-based approach reduces the dimension of the feature space, accelerating the learning process and improving the precision of 1-star versus 5-star rating classification. A two-step filter approach was applied. We selected the top-ranked features based on tf-idf values and subsequently applied a maximum relevance minimum redundancy (MRMR) approach based on a mutual information statistical measure (Peng, Long, & Ding, 2005). First, terms that could be strongly associated with other terms but weakly associated with star-ratings (i.e., lowly predictive) were discarded. Second, we applied a heuristic approach to determine an appropriate number of features by iteratively testing the naive Bayes algorithm. Naive Bayes, a simple but efficient classifier, has been shown to work well in many domains. It offers a suitable classifier for large datasets (Abellán & Castellano, 2017). Assuming that our research aim was to enhance classification rates, the features were selected according to the rates of change of their Matthews correlation coefficient (MCC) (cf. Matthews, 1975; see Equation 1). The MCC provided a coefficient of the correlation between the observed and predicted binary classifications by simultaneously considering true and false positives and negatives.

In short, our approach has three notable features. First, we initially sorted the features (terms) according to their relevance in determining class labels based on tf-idf values and

MRMR metrics. Second, each naive Bayes model was based on increasing the number of sorted features each time from 1 to 424 features. For each cumulative number *i* of terms, a naive Bayes classifier was executed. We employed the 10-fold cross validation (i.e., the training data were randomly split into 10 mutually exclusive folds). This trained the classifier on nine sets and tested it using the one remaining set. We estimated the MCC values applied to increasing the number of candidate features and selected the model that gave the maximum MCC value. The comparison of feature selection obtained by estimating MCC values is summarized in Figure 1. Figure 2 also shows the results for average accuracy, precision, recall, F1 score, Kappa, and area under the curve (AUC), as we varied the number of features. Third, rates *i* of MCC change (between term *i* and term *i-1*) were calculated as $MCC_i - MCC_{i-1}$. Our approach omitted features whose rates of MCC change were negative.

Equation 1. Matthews correlation coefficient

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP + FP)\ x\ (TP + FN)\ x\ (TN + FP)\ x\ (TN + FN)}}$$

TN: true negative; TP: true positive; FN: false negative; FP: false positive.

4. Results

Although satisfaction classification was likely to have clear feature dependencies, our best model (based on the MRMR metric) had an MCC measure of 0.734. The optimal number of features in the naive Bayes model was 209. Figure 3 shows the top-ranked features for the best model.

Insert Figure 1.

Insert Figure 2.

Insert Figure 3.

The resulting classifier was also evaluated using additional standard metrics (see Figure 4). For our prediction based on hotel ratings, the recall (based on false negatives) and precision (based on false positives) of the classifier were equally important. We assessed both precision and recall in their combined form using the F1 score. This score is used extensively in text classification. The F1 score, precision, and recall are highly informative evaluation metrics for binary classifiers (cf. Saito & Rehmsmeier, 2015). Having established the reference class as a 1-star rating, we observed an F1 score equal to 0.85. The precision score was equal to 0.89. It was conceptualized as 89% correct predictions (TP) of all classified ratings (TP: 1,280 plus FP: 163). Recall was equal to 0.81 and was conceptualized as 81% correct predictions (TP) of relevant ratings (TP: 1,280 plus FN: 295). Having established the reference class as a 5-star rating, we observed an F1 score equal to 0.88. The precision value was equal to 0.85 (TP: 1,629 plus FP: 295). Recall was equal to 0.91 (TP: 1,629 plus FN: 163). In short, more negative reviews (1-star scores) were wrongly recognized as positive reviews (5-star scores). Figure 4 evidences a higher recall of positive reviews and a higher precision of negative reviews.

The Kappa metric was used to compare observed accuracy with expected accuracy (cf. Cohen, 1960). Kappa values of 0, 1, > 0, and < 0 indicate chance, perfect, above chance, and below chance agreement, respectively (Cohen, 1960). The Kappa value was 0.73. The AUC of the receiver operating characteristics (ROC) (cf. Fawcett, 2006) is associated with the classifier's ability to avoid false classification. It is calculated using the table of confusion. The ROC is associated with the trade-off between a classifier's true positive rate (sensitivity, or TPR, on the y-axis) and false positive rate (specificity, or FPR, on the x-axis). An acceptable classification should have an FPR value close to 0 and a TPR value close to 1. The AUC is 0.5 for random and 1.0 for perfect classifiers. The AUC value was 0.86.

Insert Figure 4.

Finally, although counting the tf-idf and MRMR values is useful in knowledge extraction, the graphical distribution and the association of terms in relation to other terms and community structures are also necessary to understand the contexts in which guests' opinions are used (cf. Newman, 2006). We proposed the identification of communities of terms with similar connectivity patterns (Fortunato, 2010) by maximizing the modularity measure (cf. Brandes *et al.*, 2008). We applied the optimal detection algorithms proposed by the igraph package (Csardi & Nepusz, 2006).

The network graph was based on a Pearson correlation coefficient greater than 0.10 and p-values less than 0.001. The network graph had 72 nodes and 148 edges. Although the variables were non-metric (i.e., DTM), they were considered suitable because the terms were correlated with each other (Hair, Black, Barry, Babin, & Anderson, 2009). Terms with the highest total degree centrality were "game" (degree centrality = 7), "secur" (degree centrality = 6), "stain" (degree centrality = 5), "bet" (degree centrality = 5), and "rude" (degree centrality = 5). The most frequent associations were "kid" and "circus" (frequency = 2,345), "love" and "staff" (frequency = 2,190), "staff" and "secur" (frequency = 1,723), and "rude" and "staff" (frequency = 1,565).

By applying the modularity measure over all possible partitions, we observed a modularity measure value of 0.86. Empirically, a value greater than 0.3 is a good indicator of the significant community structure of a network (Clauset, Newman, & Moore, 2004). Community detection was employed to examine the underlying semantic structure and further reduce the number of terms into meaningful groupings that were easier to interpret. Overall, 19 communities were detected by maximizing the modularity index (see Figure 5). These groups had 72 nodes, and 26% of communities accounted for 53% of all nodes. Our graph plots the features associated with hotels' tangible and intangible features that belong to different contexts of meaning associated with hospitality services. The highest degree nodes were linked to competitive advantage in entertainment driven by tourism and gambling pleasure, hotel rooms that offered the fundamental functional benefits sought by guests (e.g., a place to sleep), and staff's delivery of the service to provide a satisfactory hospitality experience. Figure 5 reflects the number of connections (degrees) that each term had with other terms. Terms that occurred together (more degrees) in reviews are shown with larger labels. Likewise, Figure 5 plots associations between terms based on edge betweenness centrality as the number of the shortest paths that go through an edge in a graph.

Insert Figure 5.

5. Discussion

5.1. Research implications

One of the main aims of this study was to gain a deeper understanding of how to classify guests' satisfaction in the hospitality industry. To pursue this aim, we explored the semantic space that represents guests' reporting of hospitality experiences by applying supervised learning and naive Bayes classification. First, although text reviews have been widely studied in the literature, there has been scarce research on knowledge extraction from text reviews and their influence on non-economic satisfaction. Online reviews are still an emerging phenomenon. However, their significance in different sectors means that they have become a prominent research area in management research. Second, techniques such as supervised learning are relevant in marketing research, yet such techniques have not been systematically studied.

Although a growing body of research analyzes the content of non-structured reviews and explores classification algorithms, performing joint analysis of these issues to understand why guests rate hotels positively or negatively remains an under-explored area.

According to certain scholars, classification is one of the most common learning models in data mining (e.g., Ahmed, 2004; Berry & Linoff, 2004; Carrier & Povel, 2003). The aim is to build a model to predict guests' behavior by classifying data into predefined classes based on certain criteria (cf. Ngai, Xiu, & Chau, 2009, p. 2595). By using techniques for Natural Language Processing, we confirm that naive Bayes offers an effective and efficient classification algorithm in polarity analysis of reviews in marketing. The results should be more reliable and accurate than prior statistical results obtained from traditional satisfaction surveys based on small data samples. Despite its naive design and apparently over-simplified assumptions, these (negative) concerns has a small impact on performance.

5.2. Managerial implications

This study was based on a large, unstructured, complex dataset. The results of the study can help hotels allocate resources to the areas that matter to guests, identify guests' perceptions, explore and assess the influence of these perceptions on meeting guests' needs, and identify key communities of terms around which guests positively or negatively evaluate hospitality experiences.

If a hotel company seeks to satisfy guests, managers must consider classifying features of the hotel when allocating marketing resources. Attributes or benefits based on staff experience (i.e., intangible cues related to the immaterial nature of services, such as interacting with friendly employees) and professionalism were identified as an essential community of terms. Likewise, a core (tangible and experiential) community of terms based on poor design and upkeep of rooms, furnishings, and experiential aspects such as cleanliness or dirtiness is an essential driver. For hotel guests, attributes or benefits based on staff experience and professionalism (i.e., interpersonal aspects of the service such as interacting with employees who are rude or who deliver a poor service) are important elements when evaluating hotel quality. Nevertheless, core or functional reviews (e.g., poor room design or upkeep) determined by tangibility should also be addressed. Moreover, events in Las Vegas such as gambling-based attractions are key marketing propositions in promoting hotels and hospitality experiences.

Positive reviews by guests can have a significant effect on other guests' decision-making processes. Conversely, negative reviews can easily undermine the loyalty of potential customers and cause them to focus on negative eWOM. The satisfaction (sentiment) classifier was associated with transforming UGC into quantitative data. These data were used to score comments as positive or negative and thereby shape important decisions for the future of hospitality organizations. In this study, we build a classification strategy that can be applied to businesses such as hotels, movie theaters, and so forth.

Our model helps hotel managers understand guests' satisfaction in terms of star ratings. Our model also helps process huge amounts of review data by using a supervised machine learning approach. Although guests' reviews are valuable to other guests and are essential in improving the service provided by hotels, the massive number of reviews makes it difficult for managers to obtain an intuitive hotel evaluation. The primary managerial contribution of this research is to provide an approach for summarizing hospitality service reviews by removing data considered as non-informative for classification.

5.3. Limitations

This study has several limitations. First, it did not examine infrequent terms in the long tail of the distribution. A second limitation relates to self-selection bias when guests post their reviews. Third, the study focused on reviews of urban hotels in Las Vegas. These reviews might reflect some bias of visitors to Las Vegas. The location-based environment could influence the relative importance of features. Moreover, average star scores may be affected by a culture-conditioned response style (Dolnicar & Grün, 2007). Fourth, this research relies only on reviews from Yelp. Our dataset could thus have some invisible bias.

5.4. Suggestions for future research

Improving the interactive features and service-related information available on online services does not necessarily guarantee customer loyalty. Although customer reviews offer star-based ratings (an overall satisfaction measure ranging from negative 1-star reviews to positive 5-star reviews), these star ratings do not provide detailed information on customer loyalty. Accordingly, future studies should propose an evaluation of the additional effect of affective trust (Singh & Sirdeshmukh, 2000). Likewise, although a growing number of studies focus on the mechanisms through which UGC generates user satisfaction or trust (e.g., Sánchez-Franco et al., 2018; Sparks & Browning, 2011), future research should also assess the psychological sentiments through which attitudes based on the continuation of relationships are formed (Wetzels, de Ruyter, & van Birgelen, 1998). Finally, future studies should analyze gender-based differences because men and women cognitively structure hotel experiences using different criteria (Sánchez-Franco et al., 2016).

6. Conclusions

This study describes a supervised classification approach to identify the most relevant terms and their influence on hotel ratings. Yelp was used as a case study. The study focused on Las Vegas hotel reviews that were voluntarily posted by users. Given the increasing need for information organization and knowledge discovery from text data, we applied the MRMR approach based on the mutual information criterion to mine the complex features from guests' reviews and systematically perform feature selection. The aim of extraction and selection was to identify a subset of terms occurring in the training set and use this subset as features in text classification.

We applied the naive Bayes approach to solve the problem of dealing with huge numbers of service reviews, enable automatic detection of guests' satisfaction with a specific hospitality service (i.e., positive 5-star reviews or negative 1-star reviews), and consider brand building, product development, and quality assurance. The proposed opinion mining system enabled a binary classification of guests' reviews with a high MCC (> 0.73) by discarding terms that were strongly associated with other terms but weakly associated with star-ratings (i.e., lowly predictive). Most reviews were correctly classified as positive or negative (i.e., AUC > 0.86). Our proposed system was thus able to identify the polarity of guests' opinions. Despite being a naive classifier and despite its unrealistic independence assumption, the naive Bayes algorithm was able to classify reviews of Las Vegas hotels with high precision and recall (F1 score > 0.84 for both reference values) and with a low computational cost.

The results show that using term unigrams is appropriate. The results indicate that the system is fast, scalable, and accurate at analyzing guests' reviews and determining guests' opinions.

References

Abellán, J., & Castellano, J. G. (2017). Improving the Naive Bayes Classifier via a Quick Variable Selection Method Using Maximum of Entropy. *Entropy*, 19(6), 247.

Ahmed, S. R. (2004). Applications of data mining in retail business. *Information Technology: Coding and Computing*, 2, 455–459.

Berry, M. J. A., & Linoff, G. S. (2004). Data mining techniques second edition – for marketing, sales, and customer relationship management. New York, NY: John Wiley & Sons.

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15(3), 31–40.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. N. Srivastava & M. Sahami (Eds.), *Text Mining: Classification, Clustering, and Applications* (pp. 71–94). London: Taylor and Francis.

Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., & Wagner, D. (2008). On modularity clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2), 172–188.

Carrier, C. G., & Povel, O. (2003). Characterising data mining software. *Intelligent Data Analysis*, 7, 181–192.

Chaves, M. S., Gomes, R., & Pedron, C. (2012). Analysing reviews in the Web 2.0: small and medium hotels in Portugal. *Tourism Management*, 33(5), 1286–1287.

Chong, A. Y. L., Khong, K. W., Ma, T., McCabe, S., & Wang, Y. (2018). Analyzing key influences of tourists' acceptance of online reviews in travel decisions. *Internet Research*, 28(3), 564–586.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review*, E 70, 066111.1–066111.6.

Clemons, E., Gao, G., & Hitt, L. (2006). When online reviews meet hyper differentiation: a study of the craft beer industry. *Journal of Management Information Systems*, 23, 149–171.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems. Available at http://igraph.org.

Dolnicar, S., & Grün, B. (2007). Assessing analytical robustness in cross-cultural comparisons. *International Journal of Tourism, Culture, and Hospitality Research*, 1(2), 140–160.

Dorsch, M. J., Swanson, S. R., & Kelley, S. W. (1998). The role of relationship quality in the stratification of vendors as perceived by customers. *Journal of the Academy of Marketing Science*, 26(2), 128–142.

Douglass, W., & Raiento, P. (2004). The tradition of invention: conceiving Las Vegas. *Annals of Tourism Research*, 31(1), 7–23.

Duan, W., Gu, B., & Whinston, A. B. (2005). Do online reviews matter? An Empirical Investigation of Panel Data. *Decision Support Systems*, 45, 1007–1016.

Duda, R., & Hart, P. (1974). Pattern Classification and Scene Analysis. *IEEE Transactions on Automatic Control*, 19, 462–463.

Fawcett T. (2006). An introduction to ROC analysis. *Pattern Recognition Lett*ers, 27, 861–874.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *The Journal of Machine Learning Research*, 3, 1289–1305.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174.

Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3), 291–304.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23, 1498–1512.

Gretzel, U., & Yoo, K. H. (2008). Use and Impact of Online Travel Reviews. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism* (pp. 35–46). Vienna: Springer.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.

Hair, J. F., Black, W. C., Barry, J., Babin, B. J., & Anderson, R. E. (2009). Multivariate data analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall.

He, W., Tian, X., Tao, R., Zhang, W., Yan, G., & Akula, V. (2017). Application of social media analytics: a case of analyzing online hotel reviews. *Online Information Review,* 41(7), 921–935.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52.

Hornik, K., Mair, P., Rauch, J., Geiger, W., Buchta, C., & Feinerer, I. (2013). The textcat Package for n-Gram Based Text Categorization in R. *Journal of Statistical Software,* 52(6), 1–17.

Huang, J. H., & Chen, Y. F. (2006). Herding in online product choice. *Psychology & Marketing*, 23(5), 413–428.

Krippendorff, K. (2012). Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 2–26.

Langley, P., Iba, W., & Thompson, K. (1992). An analysis of Bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 223–228). AAAI Press and MIT Press.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468.

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151.

Manes, E., & Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: An empirical investigation of online hotel booking. *Journal of Business Research*, 85, 185–196.

Marchiori, E., & Cantoni, L. (2015). The role of prior experience in the perception of a tourism destination in user-generated content. *Journal of Destination Marketing & Management*, 4, 194–201.

Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta - Protein Structure*, 405, 442–451.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185–200.

Newman, M. E. J. (2006). Modularity and community structure in networks. *PNAS*, 103(23), 8577– 8582.

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592–2602.

Niu, R. H., & Fan, Y. (2018). An exploratory study of online review management in hospitality services. *Journal of Service Theory and Practice*, 28(1), 79–98.

Oliver, R. (1999). Whence consumer loyalty? *Journal of Marketing*, 63(4), 33–44.

Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and their implications for destination marketing. *Journal of Travel Research*, 46(1), 35–45.

Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing*, 87(4), 598–612.

Park, D. M., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 11, 125–148.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238.

Prabu, K. (2014, February 12). Vast majority of TripAdvisor users read at least 6–12 reviews before choosing hotel. *Tnooz*. Retrieved from http://www.tnooz.com/article/tripadvisor-online-review-insights-phocuswright-study/.

Racherla, P., Connolly, D. J., & Christodoulidou, N. (2013). What determines consumers' ratings of service providers? An exploratory study of online traveler reviews. *Journal of Hospitality Marketing & Management*, 22(2), 135–161.

Raguseo, E., Neirotti, P., & Paolucci, E. (2017). How small hotels can drive value their way in infomediation. The case of 'Italian hotels vs. OTAs and TripAdvisor'. *Information & Management*, 54(6), 745–756.

Raychaudhuri, S., Schütze, H., & Altman, R. B. (2002). Using text analysis to identify functionally coherent gene groups. *Genome Research*, 12(10), 1582–1590.

Ren, G., & Hong, T. (2017). Investigating online destination images using a topic-based sentiment analysis approach. *Sustainability*, 9(10), 1765.

Reza Jalilvand, M., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet Research*, 22(5), 591–612.

Rowley, R. J. (2015). Multidimensional community and the Las Vegas experience. *GeoJournal,* 80(3), 393–410.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* 10.3, e0118432.

Salton, G., & McGill M. H. (1986). *Introduction to Modern Information Retrieval*. New York, NY: McGraw-Hill.

Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2016). Online Customer Service Reviews in Urban Hotels: a Data Mining Approach. *Psychology & Marketing*, 33(12), 1174–1186.

Sánchez-Franco, M. J., Roldán, J. L., & Cepeda, G. (2018). Understanding relationship quality in hospitality services: a study based on text analytics and Partial Least Squares. *Internet Research*, in press.

Serra Cantallops, A., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, 41–51.

Singh, J., & Sirdeshmukh, D. (2000). Agency and trust mechanisms in consumer satisfaction and loyalty judgments. *Journal of Academy of Marketing Scienc*e, 28(1), 150–67.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32, 1310–1323.

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: the effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, 53, 74–85.

Sun, T., Youn, S., Wu, G., & Kuntaraporn, M. (2006). Online word-of-mouth: an exploration of its antecedents and consequences. *Journal of Computer-Mediated Communication*, 11, 1104–1127.

Tang, B., & He, H. (2015). ENN: Extended nearest neighbor method for pattern recognition. *IEEE Computational Intelligence Magazine*, 10(3), 52–60.

Tang, B. (2016). Toward Optimal Feature Selection in Naive Bayes for Text Categorization. IEEE Transactions on Knowledge and Data Engineering, 28(9), 2508–2521.

Tsao, W. C., Hsieh, M. T., Shih, L. W., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management*, 46, 99–111.

Vermeulen, I. E., & Seegers, D. (2009) Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127.

Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. Proceedings of the 23rd international conference on Machine learning. *Proceedings of the 23rd International Conference on Machine Learning*, 977–984.

Wetzels, M., de Ruyter, K., & van Birgelen, M. (1998). Marketing service relationships: The role of commitment. *Journal of Business & Industrial Marketing*, 13(4/5), 406–423.

Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: a meta-analysis of effect size. *Journal of the Academy of Marketing Science*, 21(2), 101–112.

Wu, L., Shen, H., Fan, A., & Mattila, A. S. (2017). The impact of language style on consumers reactions to online reviews. *Tourism Management*, 59, 590–596.

Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: a text mining approach. *International Journal of Hospitality Management*, 55, 57–69.

Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. *Proceedings of the Fourteenth International Conference on Machine Learning*, 97, 412–420.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of eWord-of-Mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634–639.

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management*, 26(1), 45–56.

Zhang, Y., Rong, J., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1), 43–52.

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: an exploration of travelers' hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*, 23(7), 972–981.

Zhu, F., & Zhang, X. M. (2010). Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *Journal of Marketing*, 74, 133–148.

Figure 1:

Evolution (by number of features) of the Matthews correlation coefficient (MCC)

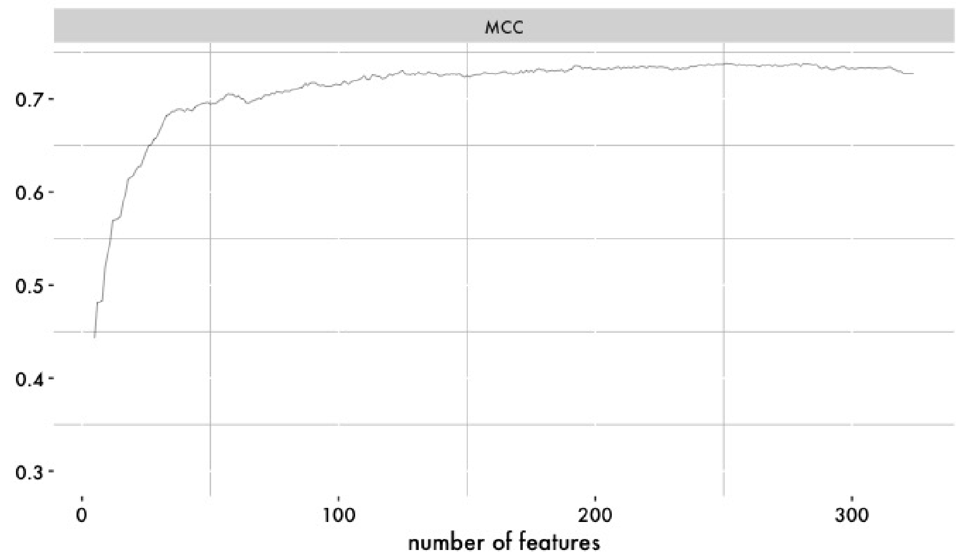


Figure 2:

Evolution (by number of features) of the classification-based metrics

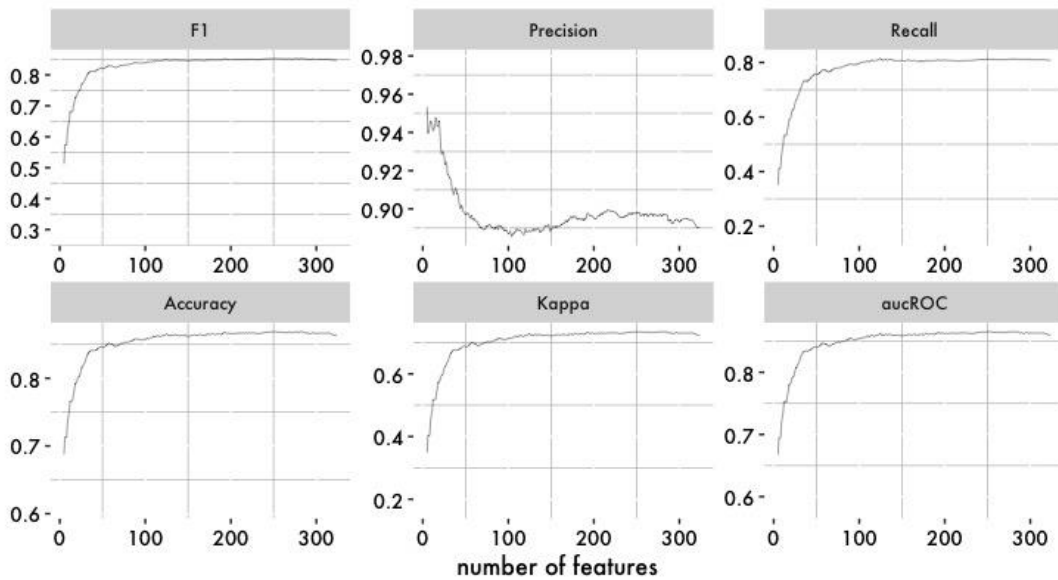


Figure 3:

Top-selected features based on positive rates of change of the Matthews correlation coefficient (MCC)

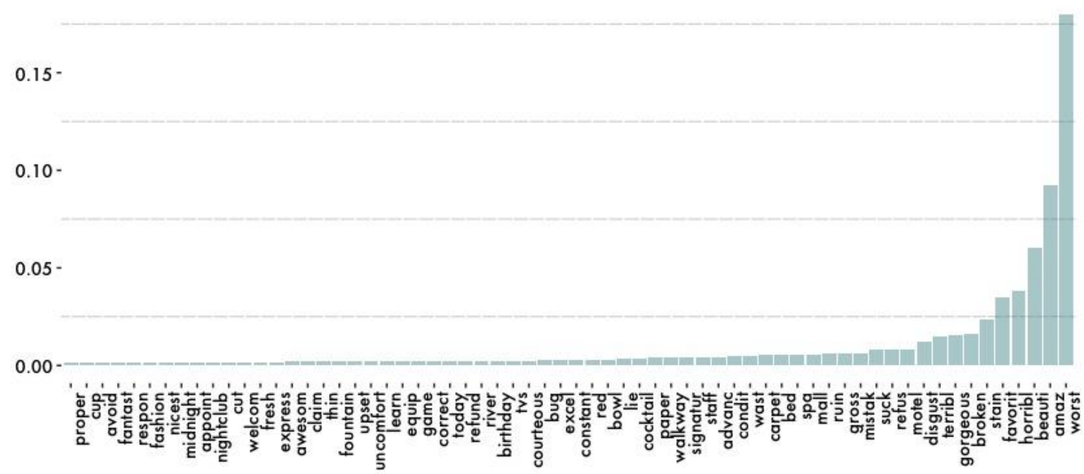


Figure 4:

Table of confusion for performance evaluation

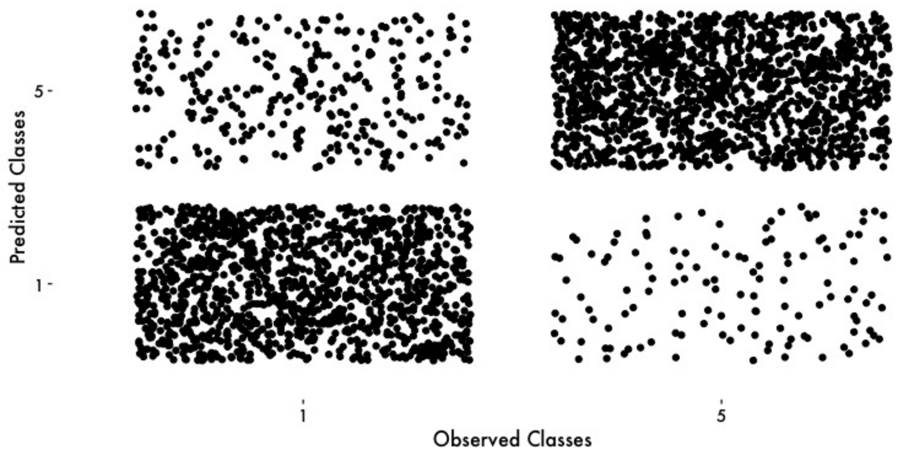
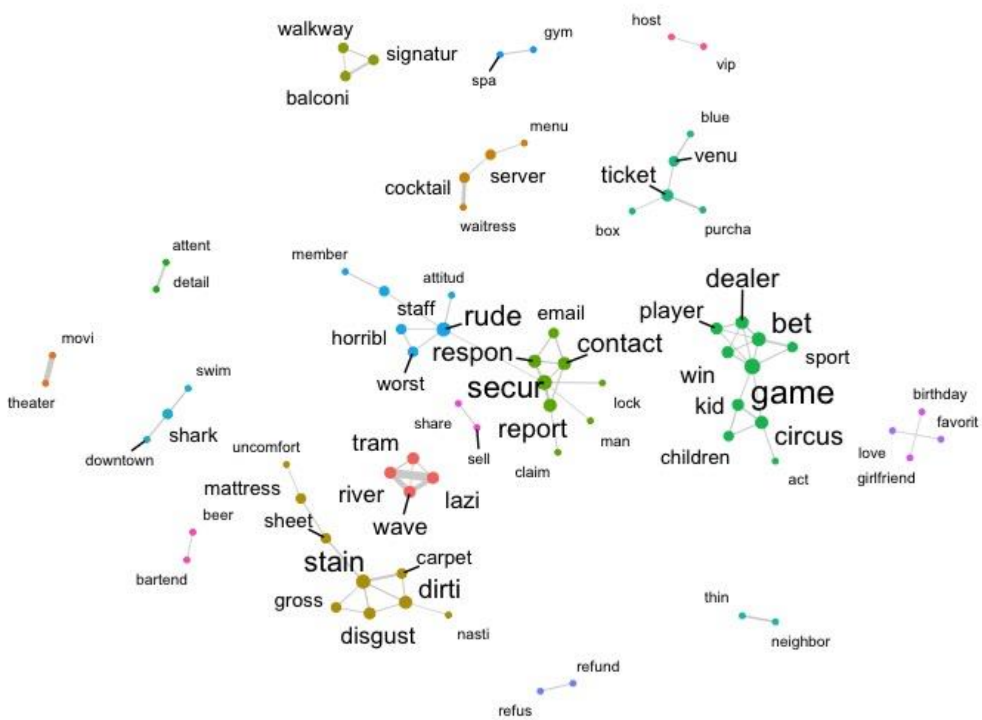


Figure 5:

19 communities: Results of modularity optimization



Note: The color of each node indicates membership to a modularity community.