ASERC Journal of Socio-Economic Studies

Journal homepage: www.ajses.az Volume5, Number 1, (2022) Pages 17-33

# RESTAURATEUR'S DILEMMA: SOLVED

**Fatima Huseynova**

*Goethe University Frankfurt,*
*Frankfurt am Main,*
*Germany*

**ABSTRACT**

The rise of peer-to-peer platforms leads customers to base their opinion on electronic word of mouth more than before, which necessitates the update of factors affecting check-in behavior of customers. Though many industries have been part of this tendency, hospitality industry comes in sight most especially. In this article, the author defines the strongest influences affecting consumer decision-making process when opting for a restaurant. By processing Yelp open-source dataset and applying 9 machine learning models, it becomes clear that Random Forest model is the best tool to predict check-in behavior of diners. On top of that, the number of photos, number of friends are the best electronic factors, while trendiness, kid-friendliness and timing are the most contributing traditional aspects.

**Keywords:** Marketing, e-WOM, Yelp, decision making, checkin behavior

A
S
E
R
C

## INTRODUCTION

It is widely known that in the last decade customers in the hospitality industry refer to peer-to-peer (P2P) platforms (such as Yelp, TripAdvisor, Angie's List, etc.) more than before. The reason behind this tendency is the limited time resources and limitless information has been made available on the Internet, which inevitably direct customers to minimize their purchase-related cognitive workload (Chen and Xie, 2008; Mostafa, 2013; Hu et al., 2016; Kangale et al., 2016; Sharma and Klein, 2016; Netzer et al., 2012; Jin et al., 2015). By turning to ratings and reviews posted by other users, consumers not only make themselves familiarized with product features and functions but also get to choose certain goods and services perfectly aligned with their taste and preferences (Fang et al., 2016; Jiang et al., 2015; Zhou et al., 2015; Sameti et al., 2016; Salehan and Kim, 2016). Therefore, electronic word-of-mouth (e-WOM) services, especially, yelp.com could bear profound insights for restaurateurs, who attempt to understand factors that influence the patronage behavior of the consumers.

Fortunately, Yelp has an open data source for students who strive for learning Machine Learning (ML) techniques for academic, professional, and personal purposes. The dataset contains (including restaurants) 160,585 different businesses, 8,635,403 customer reviews, 200,000 pictures, over 1.2 million attributes like working hours, availability of parking space, and ambiance information attached to them for 8 metropolitan areas. Analyzing this huge dataset by using ML models is very crucial for predicting customers' check-in behavior for future records and synchronizing planned marketing strategies and campaigns by preferences of the target customers. On the other side, the ongoing competitive state of market presses for a detail-oriented approach towards information processing, which underlies the significant effect of reviews on making a purchase (Gursoy, 2019). Previous research also suggests that data mining can result in the restaurant's improved service quality and customer satisfaction (Gao et al., 2018).

Throughout this paper, I dwell on the influences affecting foot traffic of the restaurants via comparing 9 well-recognized ML prediction models, including Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Artificial Neural Network (ANN), Tree, Random Forest (RF), Boosting, Bagging. For this, I have developed 39 Key Performance Indicators (KPI-s). 31 of these KPI-s are internal data variables, such as cumulative number of reviews written for a particular restaurant, its ambiance (e.g., noise level and style), availability of credit card payment, etc. derived from Yelp open data resource, and 8 are external data variables derived from weather data. After processing these relevant inputs, I have dealt with data cleansing of the dataset to prepare it for analysis. During the analysis, I came to know that among these models, Random Forest yields the best result for predicting checkin behavior of the customers in terms of its TDL and GINI coefficients and hit rate, which I will be breaking down in the upcoming sections.

The paper continues with the literature review of the previous research done on the restaurant choice of diners. I divided this section into four parts: influences that do not contain e-WOM by nature, which I call traditional factors (1); Yelp as an example of P2P platforms (2); influences that evolve around UGC via e-WOM, which I call e-factors (3); machine learning and hospitality industry (4). Afterward, within the framework of this study, I introduce the data collection, research method, results, and conclusion sections.

# 1. LTERATURE REVIEW AND BACKGROUND

## *1.1 Traditional factors affecting checkin behavior*

Everyone's having her/his own subjective criteria for choosing dining places has led researchers to generalize these decision-making criteria at maximum. As a result, during the literature review, I often observed that different researchers had ranked the relevancy of these criteria differently. For instance, according to Ha and Jang et al. (2010), the most relevant and direct influence on customers is food and beverage quality, which consists of 6 elements of food freshness, portion size, menu variety, food presentation, food taste, and food temperature (Hwang and Ok, 2013).

However, demographics play an important role; in their experimental study, Moschis et al. (2003) stated that whereas for the age 54 and younger, the most influential factor is socialization opportunities and locational ease of the eatery (in terms of its distance from their home and work), for the age 55 and older, offering special offers over a certain age (along with being a comfortable place to socialize) is the most important factor. On top of that, what can not be overlooked is the atmosphere of the dining halls, as it can either enhance or diminish satisfaction level (Ha and Jang, 2010), and consequently, make customers feel relaxed or nervous (Gustafsson et al., 2006). Interestingly, a study conducted in 1988 concluded that in an upscale restaurant, people can even spend hours watching an interior/exterior design (Baker, Berry, & Parasuraman), revealing a correlation between the price level and ambiance.

This means that in fast food restaurants like KFC, where delivering food is mainly utilitarian and functional (Lin, 2004), people will be less likely to evaluate "DINESCAPE", which is a combination of six constructs: facility aesthetics, lighting, ambiance, layout, dining equipment, and employees (Ryu and Jang, 2005). It is clear that good food increases foot traffic, however, restaurateurs should also acknowledge that customers are attracted to the environments they call "cool", which is also part of their total dining experience (Garg, 2014). While perusing the foregoing studies, I also noticed that WOM, income level, payment methods, availability of parking lots are considered crucial to check-in (Moschis et al., 2015). Intriguingly, research outcomes uncovered a hidden connection between income level and payment methods for mature Americans. In that, the more affluent the patron was, the more emphasis was given to the form of payment. But lower-income groups patronized low-budget eateries, where they had to pay at the cash register by waiting in the queue (Moschis et al., 2003).

## *2.2. Yelp as an example of P2P platforms*

It can not be denied that Yelp plays an irreplaceable role in the useful review generation process by offering a reliable social media environment for restaurant users. The company was founded in 2004, by Jeremy Stoppelman and Russel Simmons. To date, it has succeeded to sustain its perseverance in the market. According to Yelp Newsroom, as of December 2021, cumulative reviews are counted at 244 million. Reviews displayed to the users are divided into three parts and follows as recommended, not recommended and removed ones operated by the algorithm of the website and to share them on Yelp, visitors are required to have a user profile (it is free). By 2021, the majority of American businesses received 4 stars (38%) by users. Though, restaurants are not the only reason people use Yelp; namely, shopping, home & local services, beauty & fitness are also among enterprise categories. I should also note that the prevailing age group among Yelp's US-originated users is 35-54, followed by the age group of 55+ (Yelp, 2022).

As Yelp does not sell any product and sources its income from the ads aired on its website by the virtue of visitor traffic, it has an impartial stand for businesses. Yelp is deemed as a safe and reliable medium (Yang et al., 2017).

### *2.3 E-factors affecting check-in behavior*

Because of their hidden quality nature, services are thought much riskier than products (Buchanan, 1977). Therefore, as a service, visiting a restaurant also could be deemed as risky conduct for consumers, which triggers them to refer to peer-to-peer platforms as credible proxies (Budescu et al., 2003). The research suggests that visual and textual reviews (also called user-generated content (UGC)) posted by other customers who have already had an experience with the product or service are considered more dependable than those generated by the company itself (Goh et al., 2013). Prior studies show that existing online reviews not only guide people to choose the proper restaurant of their preferences but also determine the tone and flow of upcoming reviews, thus, create homogeneity (Li et al., 2020).

So far, hospitality researchers have paid a noticeable amount of attention to Twitter data (Bejarano et al., 2017). The remaining studies that are based upon Yelp reviews mainly analyze mere reviews that (e.g. via the use of Sentiment Analysis) they often overlook UGC's direct impact on restaurant visitors' decision making (Kim and Tanford, 2018).

Nonetheless, whether they are positive or negative (until they are useful), online reviews are of significant importance for eliminating information asymmetry between the customers and services (Budescu et al., 2003), which in turn leads customers to make a choice. Previous researchers took multifarious approaches on elements influencing review usefulness; whilst some asserted that it is plainly influenced by review content (Ong, 2012; Pantelidis, 2010), sufficient information and plausibility of source (Hlee et al., 2018), and mode of conduct (customers benefited from sale and discounts tend to rate higher than regular customers) (Chen and Lurie, 2013), particular scholars suggested that the number of friends the user has is the fundamental element in forecasting his/her review usefulness (Lee et al., 2020). By getting into a deeper sense, Lin et al., on the other side, advocated that image-containing reviews are more promising than textual ones (2012).

### *2.4 Machine Learning and Hospitality industry*

It has been a long time since many aspects of our everyday lives have been recorded into different databases. Apart from governments, businesses and individuals also register and monitor our electronic behavior using specialized devices, which took database science to the next level (Lantz, 2021). Consequently, via ML techniques, vast data stored in these databases should be processed and turned into insights relevant for managerial decision making. As Lantz defines it, "the field of study interested in the development of computer algorithms for transforming data into intelligent action is known as machine learning". In order to utilize these ML models, for starters, the dataset should be stored, retrieved, and cleaned. After developing a model, these techniques observe and "learn" the data from previous observations, which is named training (Lantz, 2021). Based on the dataset, ML models seize similar patterns and give predictions again for similar (not identical) events. The success of these techniques depends on the nature of the data. Namely, the same model can result in different accuracy rates and coefficients.

Enterprises that rely on data intelligence rather than heuristics, which is often brought by employee experience, tend to have less unstructured, "coincidental" outcomes which help them increase their financial profit and keep business on track. Unfortunately, academia in the hospitality and tourism industry has recently commenced focusing on big data analysis via the utilization of ML models (Li et al., 2018; Mariani et al., 2018). Nevertheless, this integration could reveal future-promising intuition when supplemented with traditional research methods (Xiang et al., 2015).

## 3. RESEARCH METHODOLOGY

### *3.1 Data Collection*

The open-source Yelp dataset[4], upon which I base my analysis associates business, check-in, tips, review, image, user tables, separately. Aligned with the purpose of this study, the tables provide figures for restaurants only, which are located in 8 metropolitan areas of the USA. They are linked through business, user, review, and photo (if applicable) ID-s. Business table, for instance, contains business ID, name, location, star ratings of minimum 1.0 and maximum of 5.0, review count, working hours, categories, and business attributes, such as ambiance, noise level, what is recommended to eat, availability of alcohol, TV, parking space, reservation, delivery, and other elements. User table, on the other hand, includes user names, their ID-s, how many reviews they have written, their friends (their ID numbers, respectively), number of review reactions (useful, cool, or funny), if applicable, years when they were elite, etc. The check-in table, which has a key role in my analysis, offers particular business ID-s and dates users each time they check in there. The photo table shows the ID number of photos, captions written on them, and their physical labels (e.g., whether they are taken inside and outside).

From the remaining tables, the review table is for review texts, reactions and stars attached to them, business ID and their authors' user ID-s, whereas the tips table is for almost the same context except for denoting tips, which are by nature, shorter than reviews. After a long period of ponderation, as certain columns were in JSON format (which made them completely unstructured) and tables internally connected with one another, I retrieved these tables from PostgreSQL, where the data is stored, via running respective SQL queries to merge them. Apart from the Yelp dataset, which I define as an internal data source, I also imported external weather data from National Centers for Environmental Information[5] website to see if check-ins are interconnected with temperature. This dataset records weather conditions since 2007 and involves snowfall, maximum and minimum temperature, amount of precipitation, etc.

### *3.2 Metrics*

Defining metrics requires meticulous attention to detail and directly affects the prediction power of the model. That is, observing conflicting views in the above-mentioned academic works, I integrate traditional factors with e-factors in my KPI-s to retain a 360-degree of cause and effect relationship. I expect this approach to comprehensively respond to the modern origins of patronage behavior in the hospitality industry.

Zang et al. (2010) consider that restaurant popularity is positively correlated with reviews that are written about physical attributes, food, and service quality of that particular restaurant. In a

[4] https://www.yelp.com/dataset
[5] https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt

similar vein, results of factor analysis made by (Sun et al., 2007) conclude that interior decor, cuisine taste, price level are of the high frequency in the restaurant selection criteria. On the opposite, Book et al. assert that customer review ratings enjoy the most important role in this event and overweighs even price (2018). Therefore, I made my mind upon including all of these incompatible elements (please see table 1) in my analysis and beheld which elements take the biggest portion (also, to which extent).

**Table 1:** Internal metrics[6]

| N | Name of the metric | N | Name of the metric |
|---|---|---|---|
| 1 | business_id | 17 | business_reservation |
| 2 | business_lattitude | 18 | business_noisy |
| 3 | business_longitude | 19 | kid_friendly |
| 4 | business_price | 20 | casual |
| 5 | business_open | 21 | romantic |
| 6 | n_photo | 22 | intimate |
| 7 | cumulative_n_reviews | 23 | hipster |
| 8 | cumulative_max_friends | 24 | upscale |
| 9 | cumulative_max_elite | 25 | classy |
| 10 | cumulative_max_fans | 26 | trendy |
| 11 | cumulative_max_useful_reviews | 27 | divey |
| 12 | business_category | 28 | female |
| 13 | business_stars | 29 | male |
| 14 | business_credit_card | 30 | check_in |
| 15 | business_wifi | 31 | date |
| 16 | business_parking | | |

Table 2 represents the remainder of the metrics which I scraped from the website of NCEI. The rationale behind this step is the riveting outcomes of a prior study conducted in 2019 which concluded that weather not only affects people's moods but also how much they enjoy their visits to a cafe (Bujisic et al., 2019). Moreover, patrons are more likely to complain about the restaurant service through e-WOM, if they find the weather unpleasant. Thus, I choose weather data as an external input to see if weather conditions influence the decision-making process of restaurant customers, similarly.

[6] Numbers are not based on metrical importance.

**Table 2:** External metrics[7]

| N | Name of the metric |
|---|---|
| 1 | PRCP |
| 2 | SNOW |
| 3 | SNWD |
| 4 | TOBS |
| 5 | TMAX |
| 6 | TMIN |
| 7 | WE |
| 8 | Quarter |

In table 2, PRCP stands for precipitation which is fall of snow, rain, sleet, etc. measured in a tenth of a millimeter. Apparently, SNOW is the amount of snowfall and SNWD is the depth of that snowfall, both measured in millimeters, whilst TOBS is the temperature at the time of observation, in a tenth of degrees Celsius. Quarter, on the other side, is slightly different in structure; it signifies 4 quarters of the year, therefore, is factored. Finally, WE is shortened version of weekends and also, an implied metric.

### *3.3 Data cleaning*

The procedure of noticing and fixing errors in the dataset via filtering, merging, and translation is called data cleaning (Hair et al., 2006; Wang et al., 2005).

As I noted before, most of the metrics are stored in JSON format, meaning that variables and observations, rather than being placed on respective columns and rows, are placed in the curly brackets and same cells, separated by commas. For instance, price level, Wi-Fi, credit card, parking space, ambiance could be counted as fair examples, for which I run some queries and put them into distinct columns. But even more unstructured is the noise level inputs of the restaurants; here in place of standardized answers like "True" and "yes", observations are leveled into quiet, average, loud, and very loud. During data inspection, I noticed that even these levels are not standardized, as they are distracted by the letter "u". For instance, when the noise level is conceived as very high, it could be written as both "u'very_loud'" and "'very_loud'" (applicable for quiet, average, and loud as well). In order to tackle this problem, I grouped all forms of very loud and loud levels under the "business_noisy" metric and called them as 1, whereas the combination of average and quiet noise levels as 0.

For aggregated metrics, including cumulative_n_reviews, cumulative_max_friends, cumulative_max_elite, cumulative_max_fans, cumulative_max_useful_reviews, I used relevant SQL queries to count for unique observations, defined their maximum values, grouped by their date and business ID-s number, and finally, joined them based on user and business ID-s to the other table variables. A similar process also applies for check-in and genders; however, for genders after aggregating distinctive user names, I used" genderdata" package of R to determine whether a particular user is male or female based on his/her name, whereas for check-in, the date

[7] All observations are recorded on a daily basis.

when the user checked in was imported and replaced with a string value of 1, if applicable. Among the weather data variables, I only derived weekend information from the temperature details of all days on the spectrum and assigned them a binary value" True" if they occurred on Saturday or Sunday.

Having gathered all 39 determinants in one table, I used the "MICE" package of R to deal with missing values. This package is a very effective tool for data cleansing because it imputes all absent datapoints with plausible values that are derived from distribution, specifically generated for each element. Assuredly, I also tested the class imbalance. As the ratio was 1318:1100, no class imbalance was detected. Though, to be sure, I applied SMOTE (synthetic over-sampling technique) to the data and compared the results. I examined no difference, thus, came to a conclusion that I can skip this step. As a last cleaning set-up, I normalized the data that made the whole dataset became ready for the analysis.

### *3.4 Prediction Method*

For the purpose of the paper, which is to predict check-in behavior of the customer and consequently, to find which algorithm works best for this goal, I am adopting nine ML algorithms: Logistic Regression, K-Nearest Neighbor, Naive Bayes, Support Vectors Machines, Artificial Neural Networks, Tree Model, Boosting, Bagging and last but not least, Random Forest model. But they are varying in their complexity level; for example, LR, NB, KNN, as well as decision trees are the simplest models which are also known as the predecessors of the data science toolkit (Lantz, 2021). SVM, ANN, however, are more tangled as they exhibit better performance at the expense of comprehensibility (also known as "black box") (Lantz, 2021). The cause for this is that these models operate on a very advanced level of mathematics which makes them hard for employing and understanding. Thanks to their working mechanism, RF, Bagging, and Boosting have gained more acceptance from the crowd (Lantz, 2021). In particular, they are considered as the most powerful supervised learning methods, as they not only handle over-fitting but also determine key subtle patterns in the data via using the diversity of the test data. Lantz asserts that one particular algorithm can not perform perfectly across all types of data (2021), therefore, I strongly believe that deploying many and various types of models will result in a more accurate proxy.

**Table 3:** Summary of the ML algorithms that are used in the analysis.

| **Model** | **Explanation** |
|---|---|
| Logistic Regression | Being part of generalized linear models family which builds a model where the relationship between explanatory variable X and dependent variable Y is displayed. Here, binary Y is assumed to obey binomial distribution (Levy et al., 2020). |
| Naive Bayes | NB takes semantics to represent, use and learn the data. It has a fair ability to make a prediction, however, often criticized for its dependencies between attributes (Yaman et al., 2019). |
| K-nearest Neighbors | Without building a model, uses the data directly for classification task. In this model, only changeable parameter is *k*, number of nearest neighbors. Its advantage is interpretability (Dreiseitl et al., 2003). |
| Support Vector Machines | Via decoding a constrained quadratic optimization problem, SVM constructs optimal separating boundaries between data blocks. In the last decade, it has gained significant interest from researchers of different fields. But results are completely dichotomous (Dreiseitl et al., 2003). |
| Artificial Neural Network | In AMM, inputs and outputs - internally connected through layers - build neuron networks. These neurons utilize weighted links to create a network and each of them correlate with input feature vectors. They can handle noisy data successfully and are also intrinsically parallel (Yaman et al., 2019). Suffers from its "black-box" nature. |
| Tree Model | This algorithm repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure. Is not a black-box model. Mainly used in medicine (Dreiseitl et al., 2003). |
| Bagging | Originally called bootstrap aggregation. Develops multiple classifiers from a succession of independent bootstrap instances (Quinlan, 2006). |
| Boosting | As bagging, this model also combines several models to create a new model. But BM handles it in a sequential, stage-wise way. It employs appropriate methods to gradually increase the focus on observations that were handled poorly by the other models before (Elith et al., 2008). |
| Random Forest | Since it has a high accuracy rate and its training time is comparably short, RF has seen considerable interest from many fields. RF is a powerful ensemble method; it randomly subsets data by bootstrapping, divides into sets and trains each set. Also, diminishes the overfitting problem (Lee et al., 2020). |

### *3.5 Assessment criteria*

To validate final model, I used Receiver Operating Characteristic Curve (ROC), confusion matrix, top-decile lift (TDL) and GINI coefficients, and accuracy rate. Parallel to this step, I changed either tuning parameters or methods of the models to see change the flow and to achieve the best prediction accuracy. Excluding LR (since both of its adopted methods have no tuning parameters), for the other models, by eliminating any kind of subjective misconception that may arise, I use automatic tuning to see which metric is decided as the most optimal. Then I compare them with their default version based on their respective measures. The accuracy rate is a traditional measure counting the percentage of correctly predicted observations in the test data group (Lemmens et al., 2006).

**Table 4:** Hyperparameter tuning

| Model | Method | Hyperparameter tuning | TDL and GINI | Accuracy rate |
|---|---|---|---|---|
| LR | #glm | no tunes | TDL=1.48<br>GINI=0.4 | 65 % |
| | #bayesglm | no tunes | TDL=1.15<br>GINI=0.14 | 65 % |
| NB | #naive_bayes | default | TDL=1.34<br>GINI=0.35 | 63 % |
| | | laplace = 0, usekernel = FALSE adjust = 1 | TDL=1.34<br>GINI=0.35 | 63 % |
| KNN | #knn | default | TDL=1.42<br>GINI=0.26 | 60 % |
| | | automatic tuning, k = 9 | TDL=1.42<br>GINI=0.26 | 60 % |
| SVM | #svmRadial | default | TDL=1.64<br>GINI=0.37 | 66 % |
| | | automatic tuning | TDL=1.53<br>GINI=0.4 | 66 % |
| ANN | #mlpML | layer1 = 5,<br>layer2 = 10,<br>layer3 = 15 | TDL=1.5<br>GINI=0.38 | 65 % |
| | | automatic tuning | TDL=1.48<br>GINI=0.38 | 63 % |
| TM | #ctree | default | TDL=1.45<br>GINI=0.34 | 64 % |
| | | automatic tuning | TDL=1.28<br>GINI=0.35 | 62 % |
| | #rpart | default | TDL=1.45<br>GINI=0.3 | 62 % |
| | | automatic tuning | TDL=1.45<br>GINI=0.24 | 65 % |
| Bagging | #treebag | default | TDL=1.5<br>GINI=0.37 | 66 % |
| | | automatic tuning | TDL=1.61<br>GINI=0.36 | 67 % |
| Boosting | #blackboost | default | TDL=1.61<br>GINI=0.4 | 64 % |
| | | automatic tuning | TDL=1.61<br>GINI=0.39 | 64 % |
| RF | #parRF | default | TDL=1.69<br>GINI=0.41 | 67 % |
| | | automatic tuning | TDL=1.64<br>GINI=0.42 | 66 % |

The logic behind not only relying on accuracy rate is that in rare situations like churn - which is of the same kind as predicting customer patronage behavior - it might be misleading (Lemmens et al., 2006). The next benchmark is TDL. It emphasizes the top 10% riskiest (those who have an actually higher probability to not visit the restaurant) customers (Lemmens et al., 2006). Whereas TDL of 1 indicates that the model is of a random model, TDL of 3 signifies that the model identifies 3 times more positive cases than a random model, and the larger the TDL, the better the model (Bose and Chen, 2009). Another important metric is the GINI coefficient, which is used for determining inequality in the predicted data. It can attain values between zero and one. Different from TDL, this metric not only focuses on the riskiest consumers but also on the less critical ones (Lemmens et al., 2006).

While being complementary with each other, TDL and GINI coefficients offer a comprehensive evaluation benchmark. Finally, I am utilizing the ROC Curve which is a two-dimensional visualization of classifier performance (Fawcett, 2004). To make a comparison among classifiers, ROC performance is usually diminished to one single scalar value, i.e. Area Under a ROC Curve (AUC). Its minimum and maximum scores are between 0 and 1.0, respectively. A classifier is considered successful if its AUC score is equal and more than 0.5 (Fawcett, 2004).

## 4 RESULTS

The outcome of the nine models[8] reveals that with a 67 per cent accuracy rate, TDL coefficient of 1.69, and GINI coefficient of 1.41, the RF model is the best tool for predicting check-in behavior of the customers in a restaurant context (please see graph 1 and 2). Following the RF model, Bagging, Boosting ensemble methods and SVM model are the next most accurate models based on their fairly similar GINI and TDL coefficients. However, due to its competitive performance, I strongly believe that the LR model (TDL of 1.48, GINI of 0.4) should not be underestimated, as it outperformed KNN, NB, TM in all indicators and is as adequate as ANN (which yielded TDL of 1.5 and GINI of 0.38). I should also note that KNN is the worst predictor among all these models with a hit rate of 60%.

I also performed hyperparameter tuning for the models (see table 4). Interestingly, it only made results better for the Bagging algorithm, via increasing TDL by 0.11, decreasing GINI by only 0.01, and improving hit rate by 1%. For KNN, NB, ANN, and Boosting it resulted in no significant changes in indicators; but for the remainder, it mostly affected TDL coefficients negatively. Therefore, besides the Bagging algorithm, I selected the default version of the models as a better option.

[8] R and SQL codes can be provided upon request.

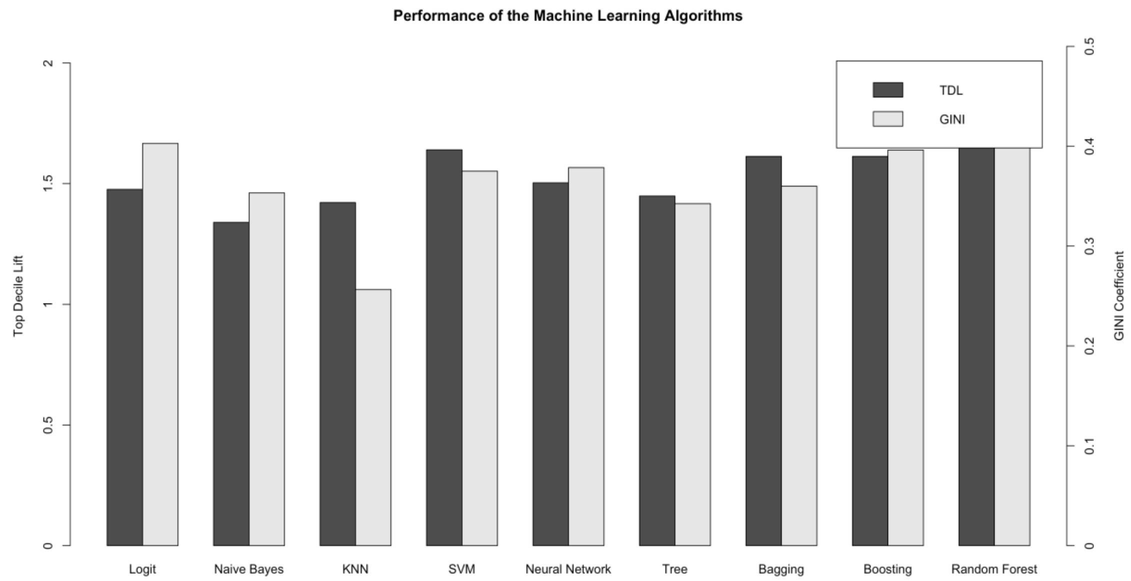



**Figure 1:** Performance of the ML algorithms

Over the analysis stage, except for LR, each of these models yielded variable importance on which a particular model had based its estimation frequently. For instance, according to RF - the top-performing model - analysis, the number of photos is the most influential factor among all 39 metrics, followed by the cumulated number of friends, review usefulness, temperature, and other KPI-s (please see Graph 3) which are mainly UGC-related. From traditional variables, weekends, trendiness, and kid-friendliness, operating with reservation characteristics of the restaurant are considered contributive. Interestingly, 3.5 stars rated businesses accounted for higher importance than 5.0 rated businesses.

**Figure 2:** AUC results of ML models.

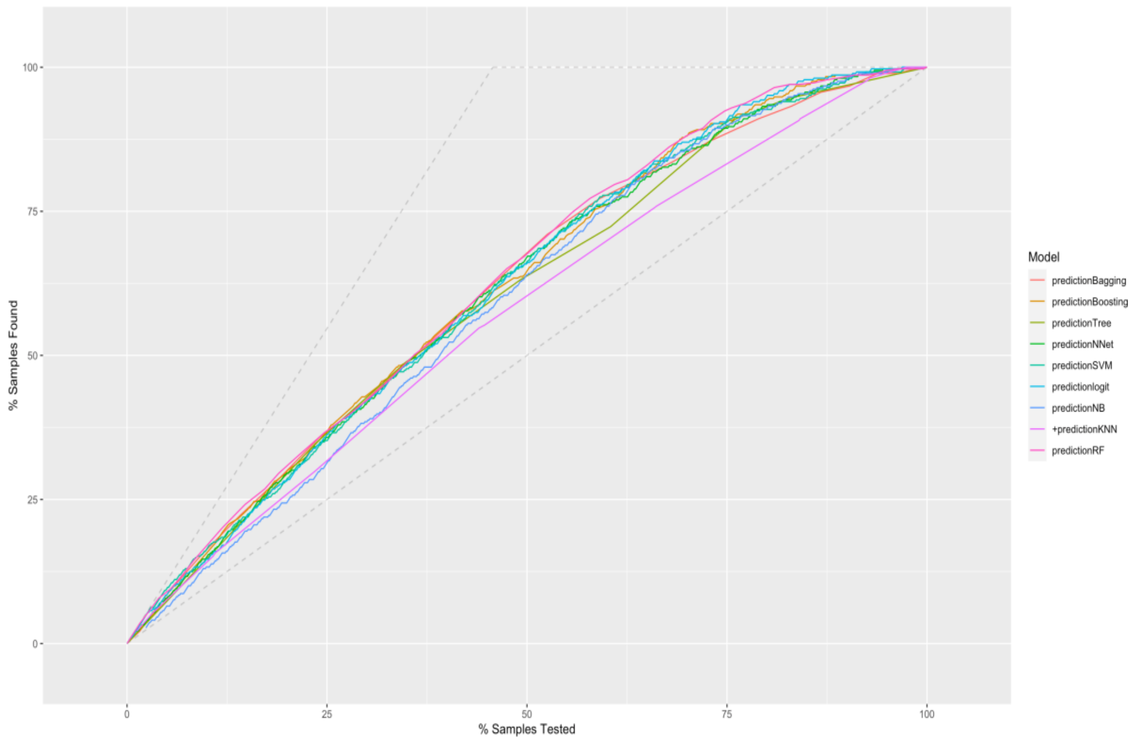
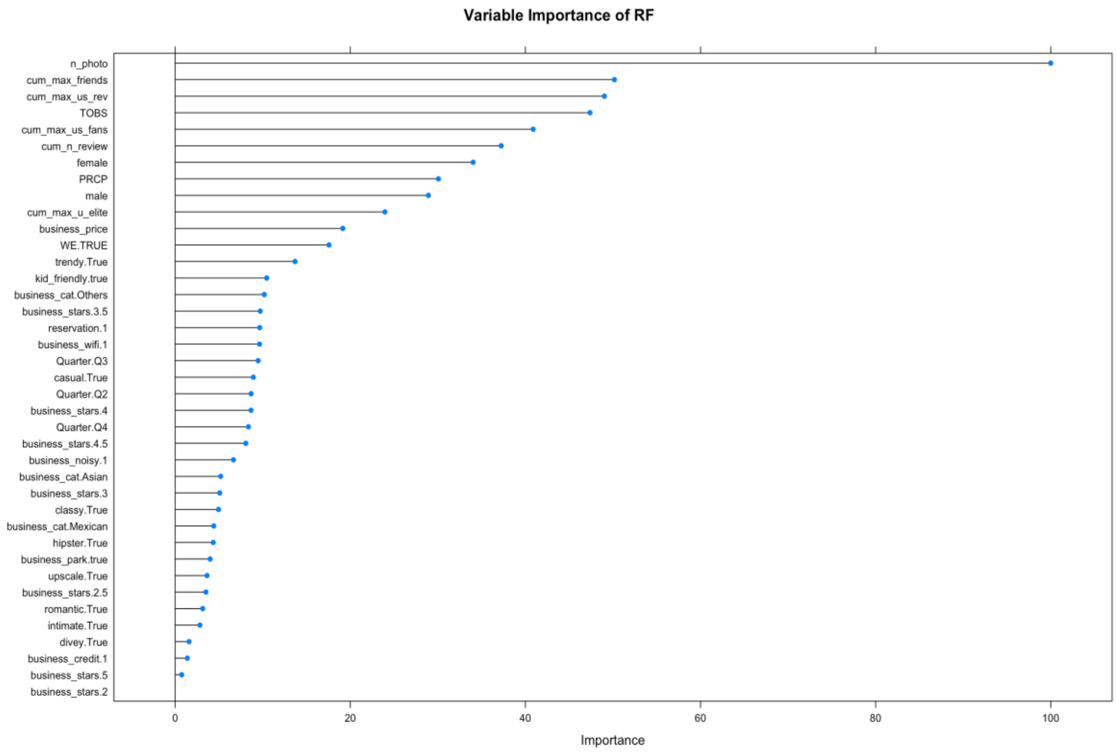


**Figure 3:** Variable importance of Random Forest model.




This can be related with the comments: only high star ratings does not mean that customer will choose a certain place. Whereas SVM, ANN, TM, KNN, NB presented almost the same variable importance as the RF model presented, Boosting ensemble algorithm benefited mostly from the gender of review writers. Bagging, on the other hand, based its prediction on the temperature at the time of observation on the highest scale. Finally, the LR algorithm, which clearly describes how particular variables affect patrons' not choosing a certain restaurant, proposed that the more it is weekend, the restaurant has Wi-Fi service, its ambiance is trendy and the temperature is high outside, the more likely customers will check in. Interestingly, to the opposite of mainstream, high noise level positively affected patrons' patronage behavior. This can be derived from the frequency of bars and clubs in these metropolitan areas, where customers dance and have a drink, rather than opt for a calm place to talk.

## CONCLUSION AND DISCUSSION

*Research implications*

The goal of this paper was to determine the list of metrics that have a significant influence on consumers' choosing where to have food. This is a striking point because marketers performing in the hospitality industry can benefit from these metrics, by addressing them in their promotional campaigns. There was a certain amount of research conducted on this topic, however, there was no consensus among the variables and most of the academia involving peer-to-peer platforms and restaurants had been deployed in Python, rather than in R. By comparing nine ML models in R, I observed that e-factors outweighed traditional factors such as weather, ambiance, additional services. E-factors, including the number of photos, friends, fans, cumulated reviews, useful reviews directly control the reader's thoughts about whether to visit this place or not. If I assume that useful reviews are those that can change the mind of a reader about a particular restaurant, then these results manifest research outcomes of Goh et al. (2013) and Lee (2020), stating that visual contents, number of reviews, and social background of the review writer - implying that they are opinion leaders - are so powerful. This is consistent with the view of Lin et al. (2012), who proposed that visual contents have more impactful force than textual contents, whereas inconsistent with Ha and Jang (2010), who assert that the strongest influence on purchase behavior comes from food and beverage quality. Outcomes also imply that the gender of the content creators is related to the check-in behavior, which approves a viewpoint suggesting that genders indeed play a role in the reader's perception of trust (Flanagin et al., 2003).

The importance of "offline" factors can not be underestimated, though. Findings approve Bujisic, et al.'s conclusions that if the temperature is higher, therefore, there is no precipitation, patrons will be positively affected. Besides, this paper offers an additional ground for 8w, who suggests that online reviews surpass the price level of the business.

Graph 3 also reveals that in its ambiance, if the business is trendy, casual, or classy; if it offers Wi-Fi and reservation service, it will attract customers. This outcome is also consistent with previous works (Baker et al., 1988; Garg, 2014; Yang et al., 2017; Sun et al., 2006; Ryu and Jang, 2007), stating that these aforementioned attributes directly affect customer's behavioral intentions.

The whole scope of this study strongly highlights that integrating Yelp data and ML techniques bears huge potential for the hospitality industry, therefore, should earn adequate credit by researchers. The hospitality industry can benefit from it, especially, if employed with future updated Yelp data covering COVID-19.

*Managerial implications*

This analysis holds considerably large implication potential for restaurateurs, who try to maximize business profits both in the short- and long-term. As the variable importance graph of the RF model describes (figure 3), Yelp-related metrics are listed as a top priority, which means that managers should put a strong emphasis on their Yelp profiles. This involves encouraging customers to post a review with the image of the plate, dealing with negative reviews by carefully addressing them, holding an event exclusively for elite members or users with wide social connections, shortly to put, for opinion leaders. Brands should also underestimate the power of social media networks, such as Instagram, Tik Tok, etc. Via these platforms, they can build a strong brand image, then, already established positive emotions can redirect this traffic to peer-to-peer platforms.

Restaurant management should successfully allocate its resources between managing the online and physical presence of the restaurant. As trendy, casual, and classy ambiance types are the most valued ones, they should take this into account and synchronize this insight with their own interior/exterior design. In my opinion, the reason for this - especially, for trendiness - is that in an era where most people share their pictures on social media, the physical appearance of the restaurant matters significantly (being "Instagrammable"). These are not the only adjustments to make, furthermore, kid-friendliness, as one of the most important variables, signals that customers also pay attention to whether a place is good for kids or not when they decide. If applicable for the restaurant category, this adjustment can lead to a sales increase.

Another signal is the weather, quarters, and weekends. Observing that customer decision making can fluctuate across these variables, restaurateurs can recognize the upcoming sales fall and offer special deals for passive times. For instance, a restaurant can prepare special "heart-warming" meals in snowy weather to balance the sales scale and effectively promote it. Finally, restaurant brands should not excessively dwell on only improving star ratings; this study reveals that even a five-star rating can fail to communicate with customers for check-in. Hence, they should rather focus on the content of the reviews.

*Limitations and future research*

Despite its notable implications, I must acknowledge that this paper has some limitations. Firstly, variables involving cumulative numbers (e.g., reviews, friends) should take time elapse into account with the help of the "yelping_since" data variable. Because this leads users with a long history to have naturally more friends than newly registered. Secondly, "business_wifi" when it gets the value 1, includes both free and paid Wi-Fi service, which can yield exaggerated results and might be not clear enough to see the relationship. Thirdly, as I have noted before, most of the ML models used in this analysis are "black-box" in nature, therefore, hard to be validated. Variable importance only returns the model's list of mostly used variables with particular per cents. It is not clear as LR model where all metrics are presented by their respective statistical significance and how they relate to the target variable. These three limitations highlight the rising urge for further research on this topic by fixing the above-mentioned biases and validating all the pre-determined factors via a meticulously designed experiment or survey.

**REFERENCES**

1. Baker, J., Berry, L. L., & Parasuraman, A. (1988). The marketing impact of branch facility design. *Journal of Retail Banking*, *10*(2), 33-42.
2. Bejarano, A., Jindal, A., & Bhargava, B. (2017). Measuring user's influence in the Yelp recommender system. *PSU Research Review*, *1*(2), 91-104.
3. Book, L. A., Tanford, S., Montgomery, R., & Love, C. (2018). Online traveler reviews as social influence: Price is no longer king. *Journal of Hospitality & Tourism Research*, *42*(3), 445-475.
4. Bose, I., & Chen, X. (2009). Hybrid models using unsupervised clustering for prediction of customer churn. *Journal of Organizational Computing and Electronic Commerce*, *19*(2), 133-151.
5. Budescu, D. V., Rantilla, A. K., Yu, H. T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178-194.
6. Bujisic, M., Bogicevic, V., Parsa, H. G., Jovanovic, V., & Sukhu, A. (2019). It's raining complaints! How weather factors drive consumer comments and word-of-mouth. *Journal of Hospitality & Tourism Research*, *43*(5), 656-681.
7. Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science*, *54*(3), 477-491.
8. Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research*, *50*(4), 463-476.
9. Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5-6), 352-359.
10. Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology*, *77*(4), 802-813.
11. Fang, J., Wen, C., George, B., & Prybutok, V. R. (2016). Consumer heterogeneity, perceived value, and repurchase decision-making in online shopping: The role of gender, age, and shopping motives. *Journal of Electronic Commerce Research*, *17*(2), 116-131.
12. Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine learning*, *31*(1), 1-38.
13. Flanagin, A. J., & Metzger, M. J. (2003). The perceived credibility of personal Web page information as influenced by the sex of the source. *Computers in human behavior*, *19*(6), 683-701.

14. Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management*, *71*, 19-32.
15. Garg, A. (2014). Mechanic clues vs. humanic clues: Students' perception towards service quality of fast food restaurants in Taylor's University Campus. *Procedia-Social and Behavioral Sciences*, *144*, 164-175.
16. Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information systems research*, *24*(1), 88-107.
17. Gursoy, D. (2019). A critical review of determinants of information search behavior and utilization of online reviews in decision making process (invited paper for 'luminaries' special issue of International Journal of Hospitality Management). *International Journal of Hospitality Management*, *76*, 53-60.
18. Ha, J., & Jang, S. S. (2010). Effects of service quality and food quality: The moderating role of atmospherics in an ethnic restaurant segment. *International journal of hospitality management*, *29*(3), 520-529.
19. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis (6th ed.*). Uppersaddle River, NJ: Pearson Prentice Hall.
20. Hlee, S., Lee, H., & Koo, C. (2018). Hospitality and tourism online review research: A systematic analysis and heuristic-systematic model. *Sustainability*, *10*(4), 1141.
21. Hwang, J., & Ok, C. (2013). The antecedents and consequence of consumer attitudes toward restaurant brands: A comparative study between casual and fine dining restaurants. *International Journal of Hospitality Management*, *32*, 121-131.
22. Jin, J., Ji, P., Liu, Y., & Lim, S. J. (2015). Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach. *Engineering Applications of Artificial Intelligence*, *41*, 115-127.
23. Kim, E. L., & Tanford, S. (2019). Simultaneous effects of multiple cues in restaurant reviews. *Journal of Services Marketing*, *33*(5), 521-531.
24. Lantz, B. (2021). Overview of Machine Learning Tools. In *The Machine Age of Customer Insight*. Emerald Publishing Limited.
25. Lee, M., Kwon, W., & Back, K. J. (2020). Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. *International Journal of Contemporary Hospitality Management*. 33(6), 2126-2128.
26. Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, *43*(2), 276-286.
27. Levy, J. J., & O'Malley, A. J. (2020). Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. *BMC medical research methodology*, *20*(1), 1-15.
28. Li, H., Meng, F., Jeong, M., & Zhang, Z. (2019). To follow others or be yourself? Social influence in online restaurant reviews. *International Journal of Contemporary Hospitality Management*, *32*(3), 1067-1087.
29. Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301-323.
30. Lin, I. Y. (2004). Evaluating a servicescape: the effect of cognition and emotion. *International Journal of hospitality management*, *23*(2), 163-178.
31. Lin, T. M., Lu, K. Y., & Wu, J. J. (2012). The effects of visual information in eWOM communication. *Journal of Research in Interactive Marketing*, *6*(1), 7-26.
32. Mariani, M., Baggio, R., Fuchs, M., & Höpken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, *30*(12), 3514-3554.
33. Moschis, G., Curasi, C. F., & Bellenger, D. (2003). Restaurant-selection preferences of mature consumers. *Cornell Hotel and Restaurant Administration Quarterly*, *44*(4), 51-60.
    Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert systems with applications*, *40*(10), 4241-4251.
34. Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, *31*(3), 521-543.
35. Ong, B. S. (2012). The perceived influence of user reviews in the hospitality industry. *Journal of Hospitality Marketing & Management*, *21*(5), 463-485.

36. Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, *51*(4), 483-491.
37. Quinlan, J. R. (1996). Bagging, boosting, and C4. 5. In *Aaai/Iaai, vol. 1* (pp. 725-730).
38. Ryu, K., & Jang, S. (2008). RETRACTED ARTICLE: Influence of restaurants' physical environments on emotion and behavioral intention. *The Service Industries Journal*, *28*(8), 1151-1165.
39. Ryu, K., & Jang, S. S. (2007). The effect of environmental perceptions on behavioral intentions through emotions: The case of upscale restaurants. *Journal of Hospitality & Tourism Research*, *31*(1), 56-72.
40. Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40.
41. Sameti, A., Khalili, H., & Sheybani, H. (2016). Analysing motivational factors influencing selection of commercial websites by consumer in the e-commerce (B2C). *International Journal of Electronic Marketing and Retailing*, *7*(1), 39-65.
42. Sharma, V. M., & Klein, A. (2016). Intention-to-participate in online group buying and consumer decision-making styles. *International Journal of Electronic Marketing and Retailing*, *7*(3), 245-271.
43. Sun, Y. H. C., & Morrison, A. M. (2007). Senior citizens and their dining-out traits: Implications for restaurants. *International Journal of Hospitality Management*, *26*(2), 376-394.
44. Wang, X., Hamilton, H. J., & Bither, Y. (2005). *An ontology-based approach to data cleaning*. Regina, SK, Canada: Department of Computer Science, University of Regina.
45. Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. *International Journal of Hospitality Management*, *44*, 120-130.
46. Yaman, E., & Subasi, A. (2019). Comparison of bagging and boosting ensemble machine learning methods for automated EMG signal classification. *BioMed research international*, *2019*.
47. Yang, S. B., Hlee, S., Lee, J., & Koo, C. (2017). An empirical examination of online restaurant reviews on Yelp. com: A dual coding theory perspective. *International Journal of Contemporary Hospitality Management*, *29*(2), 817-839.
48. Yang, S. B., Shin, S. H., Joun, Y., & Koo, C. (2017). Exploring the comparative importance of online hotel reviews' heuristic attributes in review helpfulness: a conjoint analysis approach. *Journal of Travel & Tourism Marketing*, *34*(7), 963-985.
49. Yelp, (2022). Yelp Metrics as of March 31, 2022. https://www.yelp-press.com/company/fast-facts/default.aspx (Accessed 20.05.2022)
50. Yen, B., Burke, M., Tseng, C., Ghafoor, M., Mulley, C., & Moutou, C. (2015, September). Do restaurant precincts need more parking? Differences in business perceptions and customer travel behaviour in Brisbane, Queensland, Australia. In *37th Australasian Transport Research Forum (ATRF), Sydney, Australia, 29th Sept–2nd Oct*.
51. Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, *29*(4), 694-700.
52. Zhou, F., Jianxin Jiao, R., & Linsey, J. S. (2015). Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews. *Journal of Mechanical Design*, *137*(7), 071401.