Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic

Yi Luo [a,1], Xiaowei Xu [b,*,1]

[a] *College of Business Administration, Capital University of Economics and Business, Beijing, China*
[b] *School of Business Administration, Southwestern University of Finance and Economics, Chengdu, China*

ARTICLE INFO

ABSTRACT

Online reviews remain important during the COVID-19 pandemic as they help customers make safe dining decisions. To help restaurants better understand customers' needs and sustain their business under current circumstance, this study extracts restaurant features that are cared for by customers in current circumstance. This study also introduces deep learning methods to examine customers' opinions about restaurant features and to detect reviews with mismatched ratings. By analyzing 112,412 restaurant reviews posted during January-June 2020 on Yelp.com, four frequently mentioned restaurant features (e.g., service, food, place, and experience) along with their associated sentiment scores were identified. Findings also show that deep learning algorithms (i.e., Bidirectional LSTM and Simple Embedding + Average Pooling) outperform traditional machine learning algorithms in sentiment classification and review rating prediction. This study strengthens the extant literature by empirically analyzing restaurant reviews posted during the COVID-19 pandemic and discovering suitable deep learning algorithms for different text mining tasks.

1. Introduction

In the e-commerce era, online reviews could be utilized to help customers make better decisions that improve their quality of life. The impact of online reviews on customers' decision-making processes and product sales has been well documented (Hernández-Méndez et al., 2015; Xun and Guo, 2017; Zhang et al., 2014a, 2014b, 2014c, 2014d). Online reviews are especially important for the restaurant industry since a one-star increase in Yelp rating could bring restaurant a 5%–9% increase in revenue (Luca, 2016). Research conducted by reviewtrackers.com reported that 33 % of customers will not choose a restaurant with an average 3-star review (on a 5-point scale) on a review website such as Yelp, Google, and Facebook. It further indicated that 80 % of customers tend to use a rating filter when searching for a restaurant (Bassig, 2019). Hence, if a restaurant review is not rated as appropriate, it could influence the overall rating score of a restaurant business and subsequently affected customers' decision-making process and a restaurant's revenue performance.

It is generally assumed that numeric ratings are aligned with the sentiments conveyed in the textual reviews (Hu et al., 2014). However, the star rating provided by a customer is sometimes inconsistent with the review context. Users usually write negative sentences despite reporting 4 or 5 stars on a numeric scale that ranks from 1 (Terrible) to 5 (Excellent) on TripAdvisor.com (Valdivia et al., 2019). This inconsistent pattern is found to be more salient in fake reviews than authentic reviews (Shan et al., 2018). Whether these inconsistent text-rating reviews are fake or caused by customers who accidently click the wrong rating scale while posting their reviews, they could influence the social reputation and revenues of both online providers and review websites (Antonio et al., 2018).

During the COVID-19 pandemic, it is reported that a number of businesses received 1-star reviews for being closed or spreading the virus (Chatmeter, 2020). Some diners gave low-starred reviews on Yelp and complained about the slow service or heat waves in outdoor seating areas (Kragen, 2020). These reviews during the pandemic are making it even harder for these struggling restaurants to survive. Therefore, to help restaurants that are still open during COVID-19 maintain and improve their service quality, it is important to analyze sentiments of online reviews to better understand the opinions conveyed by the diners. In addition, to help the restaurant industry fully recover its sales gradually during the rest of 2020 and survive in the long run, it is important for review websites to identify fraudulent and problematic reviews and

---

* Corresponding author.
E-mail address: xiaoweix@swufe.edu.cn (X. Xu).
[1] These authors contributed equally to this work.

improve the trustworthiness of online reviews.

Although discrepancies between textual comments and numerical ratings have received considerable attention from researchers (e.g., Lo and Yao, 2019; Zhang et al., 2014a, 2014b, 2014c, 2014d), studies on review rating prediction using machine intelligence methods are scarce in the tourism and hospitality literature with few exceptions (Antonio et al., 2018; Zheng et al., 2021). Using big data and deep learning approaches could help hospitality and tourism practitioners discover dynamics based on large volumes of data and gain new insights that were unable to be detected with traditional approaches such as survey and interviews (Alaei et al., 2019). There are even fewer studies identifying customers' opinions and sentiments towards a restaurant in the time of COVID-19. To address the aforementioned literature gap and help restaurant better assist customers during the COVID-19 pandemic, the following research questions were used to guide this study: (1) How do restaurant customers perceive their dining experience during the COVID-19 pandemic?; and (2) How do deep learning algorithm perform in improving the accuracy of sentiment classification and restaurant review rating?.

The specific objectives of the current study are as follows: (1) to conduct a sentiment analysis on restaurant reviews posted between January 1 and June 30, 2020 on Yelp; (2) to develop a deep learning technique that could automatically access the inconsistency between a numeric rating and the associated textual review content using the same Yelp dataset; (3) to find the best model respectively for review sentiment and rating prediction by making comparisons between two deep learning techniques and two conventional machine learning techniques. It is expected that the sentiment analysis of online restaurant reviews could help restaurateurs understand the key drivers of customer sentiment during the COVID-19 pandemic and further refine their products, services, and brand image. The proposed rating prediction mechanism could help customers make smarter decisions, enhance the sales volume of restaurants, and increase the utility of the review websites.

2. Literature review

2.1. The impact of COVID-19 on the restaurant industry

Compared to other industries, the restaurant industry has suffered the most significant sales and job losses since the Covid-19 outbreak began. In the U.S., the National Restaurant Association (2020) reported that the industry lost $165 billion in sales between March and July. If the pandemic lasts for six months, the chance for survival for restaurants is expected to be only 15 % (Bartik et al., 2020). It is also estimated that 13.4 million jobs could be affected in the restaurant industry (McKinsey and Company, 2020). Using data from OpenTable, it was found that sit-in guests in most states had declined by 90 % by March 18, 2020 (Dube et al., 2020) and 60 % of restaurants are permanently closed (Croft, 2020). During the COVID-19 pandemic, a number of restaurants offer takeout and delivery options to sustain business until normal operations resume. As the U.S. gradually reopened businesses in May, every state allowed restaurants to provide dine-in service with varying social distancing guidance (Sontag, 2020).

Yelp has claimed that online reviews remain as important as ever during the COVID-19 pandemic (Rubin, 2020). Timely online reviews could help potential diners gain the most up-to-date information about the way a restaurant is operating during the COVID-19 crisis. A single negative review might deter potential customers, which could make it even harder for restaurants to survive the impact of the COVID-19 pandemic. Therefore, analyzing the sentiments in hidden topics of restaurants reviews during the first six months of 2020 could give restaurateurs a general picture of how customers behaviors and attitudes evolve during a pandemic and what customers care about in order to increase their Yelp ratings. Additionally, although Yelp has encouraged its users to remain empathetic and patient with businesses (Rubin, 2020), it will make the reviews untrustworthy and biased if a customer

reads a five-star review with negative experiences. Such online reviews with inconsistency between ratings and sentiments could also give problematic businesses an unfair advantage.

2.2. The importance of online review content and rating

Electronic word-of-mouth (eWOM) refers to any communication customers make via web-based discussion or review platforms (Brown et al., 2007). Review rating and review text are two important components of eWOM. They both exert significant influences on customers' attitudes towards a product or service and their subsequent decision-making. For example, Ha et al. (2016) found that a high review rating could enhance customers' intention to choose both dine-in and take-out restaurants (Ha et al., 2016). The congruence between affective content and linguistic style properties of customer online book reviews could influence conversion rates (Ludwig et al., 2013).

In situations where both review ratings and text are present, customers are willing to make an effort to process both pieces of information in order to make a more reasonable and accurate decision (Mudambi et al., 2014). Therefore, researchers have started to explore the interplay between review rating and content. Based on a multiple equation model, Hu et al. (2014) found that ratings do not have a direct impact on product sales, but have an indirect impact through sentiments embedded in online reviews. Studies have also confirmed the existence of inconsistency between user ratings and underlying sentiments of online reviews (Shan et al., 2018; Valdivia et al., 2019). In addition to careless reviewers and manipulations of online reviews, the potential reason for rating and text misalignment might be due to the ambiguous rating systems, customers' different perceptions towards numeric ratings (Mudambi et al., 2014), and the complexity of condensing opinions into a single number (Centeno et al., 2015).

Consequently, inconsistent text-rating valences could lead to customer confusion while seeking product/service information (Geierhos et al., 2015), decrease customers' trust in reviews (Tsang and Prendergast, 2009), increase customers' cognitive processing costs, decrease customers' satisfaction with the review site, and subsequently result in a suboptimal purchase decision (Mudambi et al., 2014).

2.3. Big data analysis techniques in the tourism and hospitality industry

Big data and analytics could benefit general businesses, especially for hospitality and tourism industry where online reviews are important data source that reflects customers' experiences and evaluation of products (Xiang et al., 2017; Yallop and Seraphin, 2020). With the exponential increase in the availability of online reviews, it is difficult and costly to conduct traditional manual content analysis. Therefore, automatic multi-aspect algorithmic and machine-operated systems are in high demand in order to analyze large volumes of data effectively (Alaei et al., 2019).

Machine learning and deep learning approaches have emerged as powerful tools to analyze online reviews through their efficient computation and intelligence. Rather than making priori assumptions as traditional statistic models, learning models allow the system to study from data (Van Calster, 2019). Better than using fixed statistics, learning models provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

2.4. Studies on review rating prediction

A number of studies have performed sentiment analysis to construct a rating prediction model in the hospitality and tourism industry. For example, Asghar (2016) extracted 1,125,458 restaurant reviews from Yelp.com and built 16 different prediction models by combining four feature extraction methods with four machine learning algorithms to predict ratings. It was found that combining logistics regression with unigrams and bigrams achieved the best predictive power. Based on an

analysis of 52,264 New York restaurant reviews from Citysearch.com, Ganu et al. (2009) proposed an ad-hoc sentiment scoring technique and regression-based scoring technique to predict ratings. The K-Nearest Neighbor algorithm (KNN), a machine learning based collaborative filtering algorithm, was implemented. It was revealed that incorporating topics and sentiment embedded in a review into a regression could better predict general review score than using the numerical star ratings given by the users. Shan et al. (2018) utilized a lexicon-based approach to make rating predictions of 24,539 restaurant reviews on Yelp. Using 1,569,263 reviews from Yelp.com, Qiu et al. (2018) proposed a rating prediction model by taking into account the sentiment of the aspects and the number of positive and negative aspects in the review. Antonio et al. (2018) applied machine learning and natural language processing approaches to develop a hotel online review rating prediction model using 23,322 reviews collected from two different sources: Booking.com and Tripadvisor.com.

Although conventional machine learning techniques have been widely used in analyzing online reviews, they were "limited in their ability to process natural data in their raw form" (Lecun et al., 2015, p. 436). However, deep learning allows the computational model involving multiple processing layers (called neural networks) to automatically discover the pattern and structure behind vast amounts of data (Lecun et al., 2015). In recent years, deep learning has gained popularity due to its capability of learning high level abstractions through its hierarchical architectures (Ain et al., 2017). Several studies have applied deep learning techniques to analyze online reviews. For example, Tang et al. (2015) integrated a user-word composition vector model (UWCVM) into a deep learning framework for review rating prediction of restaurant reviews from Yelp and movie reviews from Rotten Tomatoes. Seo et al. (2017) used attention-based convolutional neural networks to predict review ratings of restaurant reviews from Yelp and product reviews from Amazon. Based on 88,882 TripAdvisor reviews of six Italian and Spanish monuments, Valdivia et al. (2019) proposed a Polarity Aggregation Model with CoreNLP to detect inconsistency between the user polarity and the automatically extracted sentiment. Martín et al. (2018) developed and compared classifiers based on convolutional neural network (CNN) and long short-term memory networks (LSTM) using online reviews of hotels on the island of Tenerife, implying LSTM recurrent neural networks performed better in predicting review scores.

In the field of tourism and hospitality discipline, most of the exiting studies have focused on either sentiment classification or rating prediction (e.g., Calheiros et al., 2017; Schuckert et al., 2015; Zheng et al., 2021). Plus, most of previous studies have focused on utilizing machine learning algorithms in analyzing online reviews except for Zheng et al. (2021) and Ma et al. (2018b). As suggested by Rafay et al. (2020), there is still room for enhancing the effectiveness of these prediction models using advanced machine learning algorithms. However, the deep learning approach, an advanced machine learning technique, is still in its infancy for tourism and hospitality research (Zheng et al., 2021). Hence, to find a superior learning method for respective sentiment and review rating prediction tasks in the restaurant industry, there is a need to compare the prediction performance of different deep learning methods with other conventional machine learning methods.

3. Method

3.1. Data collection

The target reviews were from restaurants located in Chicago, Houston, Los Angeles, New York City, and Philadelphia, which are among the top ten largest cities by population in the U.S. (White, 2020). A web scraper was built to obtain Yelp restaurants information such as, name, price, overall rating, customer's rating, and textual reviews posted between January 1 and June 30, 2020. In total 112,412 reviews were retained for data analysis.

3.2. Data cleaning

All the textual reviews were pre-processed by two major procedures: tokenization and stopwords removal (Aggarwal and Zhai, 2012; Konchady, 2006). In the first process, all the textual data were transformed into lower-case. Phrases, sentences, paragraphs, or an entire text document were divided into single word. Website links, numbers, symbols, special characters, and punctuation were removed, followed by normalization, which change words into canonical form. The last step of tokenization is stemming, which is the process of eliminating morphological changes. The stopwords ("you", "I", "we", "she", "the", "is", "a", "he", etc.) which could not provide meaningful inputs were removed to further analysis.

3.3. Data analysis

3.3.1. Feature engineering and sentiment analysis

Term Frequency-Inverse Document Frequency (TF-IDF) was applied as a topic modeling algorithm to extract features and the corresponding key words of the whole review dataset (Quan and Ren, 2014). As one of the main attributes (e.g., value) that customers care about most, a feature includes extensive key words (e.g., cheap, expensive). TF-IDF increases along with the frequency of a word occurs in the reivew.It contains two major steps as shown in the following equations adapted from Christian et al. (2016): calculating the vocabulary set and calculating a TF-IDF output for every word in a review. We first created our vocabulary using all comments from the training set. In this step we ignored the word with frequencies of less than 5 and the words with frequencies greater than 0.8 * number of documents.

$$TF - IDF_{ij} = \text{Term Frequency} * \text{Inverse Document Frequency} \quad (1)$$

$$\text{Term frequency} = \frac{\text{frequency of word}_i}{\text{length of document}_j} \quad (2)$$

$$\text{Inverse Document Frequency} = \log \frac{\text{number of document}}{\text{number of documents containing word}_i} \quad (3)$$

With the features derived from previous step, Natural Language Toolkit (NLTK) and dictionary-based sentiment analysis tools were employed to calculate sentiment scores of our training dataset. Sentiment analysis is the automated process of classifying texts according to the emotions that customers express as positive, negative, or neutral (Mouthami et al., 2013). Dictionary-based approach has been widely utilized to perform sentiment analysis (Ma et al., 2018a, 2018b; Nie et al., 2020). SentiWords covers a wide range of English words including roughly 155,000 associated with a sentiment polarity score between -1 (negative) and 1 (positive). Different from other sentiment lexicon like SentiWordNet, SentiWords assigns sentiment scores directly to words including adjectives, nouns, verbs and adverbs. Then, dictionary-based methods calculate the total sentiment by adding up the individual sentiment polarities for each word within the comment (Taboada et al., 2011).

3.3.2. Model comparison

We examined two different deep learning and machine learning models to predict the sentiment and the ratings of comments that users left on Yelp. In terms of the traditional machine learning models, the Gradient Boosting Decision Tree (GBDT) and the Random Forest classifier were applied. In terms of the deep learning methods, the two most popular methods were applied: Simple Embedding + Average Pooling and Bidirectional LSTM.

3.3.2.1. Gradient boosting decision tree classifier.
In order to obtain a stronger learner, boosting combine a series of sequentially linked weak learners (Feng et al., 2018). GBDT is a united model of sequence training

multiple weak learners called decision trees (Rao et al., 2019). High performed model was generated through adding more trees to minimize the errors of previous trees (Ke et al., 2017). Every time a new tree joined in, it better fits the initial dataset (Rao et al., 2019). Therefore, GBDT keep fitting residual errors also called negative gradients via learning from each decision treesto achieve a strong learner (Hastie et al., 2009).

3.3.2.2. Random forest classifier. To conquer the high variance and correlation of bootstrapped decision trees, random forest classifier was employed (Pal, 2005). A random forest includes of a number of less correlated decision trees. One fundamental difference that random forest has is its randomly selection of subset factors instead of choosing the most desirable divisor among the whole factors. Thus, the most reliable predictor is generated from the previous step (Pranckevičius and Marcinkevičius, 2017). Following Oshiro et al. (2012), 128 trees were applied to build our random forest model. To eliminate over-fitting, performance was examined of each tree by modifying the amount of leaves range from 2 to 16.

3.3.2.3. Bidirectional LSTM. To make up for the deficiency of gradient disappearance in RNN, Hochreiter and Schmidhuber introduced LSTM in 1997. The introduction of an adaptive gating mechanism is its principle. This mechanism determines the degree to which the previous state is maintained, then remember the extraction characteristics of the current input data, which with the advantage of sequential processing textual inputs.

In 1997, two-way LSTM was introduced by Schuster and Paliwal to expand one-way LSTM by adding the second layer of hidden layer, in which hidden connections flow in opposite time order. A Bidirectional LSTM enables researchers to see forward by applying a forward LSTM and a backward LSTM. Both process the inputs in the chronological order. Resulting the concatenation of the corresponding states of the forward and backward LSTM at a given time step.

3.3.2.4. Simple word-embedding model (SWEM). Simple word embeddings-based model (SWEM) aims to examine the raw modeling capacity of word embedding. Specifically, models that encode natural language sequences without the use of additional components are considered. Among them, calculating the element average of a given sequence of word vectors becomes the simplest strategy (Adi et al., 2016; Wieting et al., 2015).

3.4. Accuracy measures

In order to measure the accuracy of a machine/deep learning model, the train/test split strategy was used. We allocated 80 % of the crawled dataset for training, and 20 % for testing, which is a commonly used train-test split ratio (e.g., Martinez-Torres and Toral, 2019; Sánchez-Franco et al., 2019). The parameters for four algorithms applied in current study were detailed in Table 1.

The F1-measure was employed to assess and contrast the performance of applied learning models (Chen et al., 2019; De Choudhury et al., 2013). The proportion of the true positives (TP) votes and the grand total of true positives (TP) with false negatives (FN) votes is "recall". A data point designated as positive by the model is true positive, and a false negative is the opposite. The proportion of all true positives

votes and the grand total true positives with false positives (FP) votes is "precision", while false positive refers to data points wrongly identified as positive but actually negative by the model. The definitions are as shown in the following equations:

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

Evaluating the accuracy of different prediction models depends on the evaluation of the following three measures: Mean Absolute Error (MAE), Mean Square Error (MSE), and R-squared (R^2). By estimating the variance between the prediction score generated by our model γ_i and the ground-truth score provided by reviewers $\widehat{\gamma}_i$ (i.e., MSE for the overall review dataset) is presented in Eq. (7).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\gamma_i - \widehat{\gamma}_i)^2 \tag{7}$$

MAE is defined as the average of absolute error and is also a measure of the difference between two variables. It measures the average of the absolute differences of the test sample between the prediction and actual observation where all individual differences have equal weight (as shown in Eq. (8)).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\gamma_i - \widehat{\gamma}_i| \tag{8}$$

The R^2, also named as the coefficient of determination, is defined as "the proportion of variance 'explained' by the regression model makes it useful as a measure of success of prediction" (Kim et al., 2013 p.82) as shown in Eq. (9):

$$R^2 = 1 - \frac{n \times MSE}{\sum_{i=1}^{n} (\gamma_i - \widehat{\gamma}_i)^2} \tag{9}$$

Even though there is no universally accepted cutoff point for aforementioned accuracy measures like MSE, MAE, F1, it is widely acknowledged that the lower the MSE and MAE values, the higher the accuracy of prediction (e.g., Hyndman and Koehler, 2006). The best and worst values of F1-measure (precision and recall) are 1, 0 (Hackeling, 2017). In terms of R^2, the more the value of R^2 close to 1, the better the model fits the data (Redell, 2019).

4. Results

4.1. Basic information of the data and settings

Table 2 provides the data profiles of 112,412 textual reviews. Los Angeles and New York City offered the highest volumes of customer reviews out of the five cities with 32.1 % and 21.6 %, respectively. The reviews with 5 stars (52.7 %) and 4 stars (22.2 %) were dominant, while the reviews with 3 stars and below accounted for 25.1 %. Around 40 % of the total reviews were provided in the first two months of 2020, followed by a decline to 15 % in March and further dropped to 8% in April. However, for the last two months in the first half of the year,

Table 1
Parameters used for each algorithm.

Gradient Boosting Decision Tree Classifier		Simple Embedding + Average Pooling		Bidirectional LSTM		Random Forest Classifier	
alpha	0.9	embedding vector	glove.6B.100d	embedding vector	glove.6B.100d	n estimators	10
learning rate	0.3	embedding dimension	100	embedding dimension	100	max depth	10
max depth	7	batch size	64	batch size	64	min samples split	50
criterion	friedman mse	epotch	5	epotch	5	max features	0.3

Table 2
Descriptive Summary of extracted Yelp reviews.

Variable	Numbers of reviews in each category (N = 112,412)	Percentage of reviews in each category (%)
City		
Chicago	22,415	19.94
Houston	19,033	16.93
Los Angeles	36,138	32.14
New York City	24,306	21.62
Philadelphia	10,520	9.36
Month		
Jan	23,461	20.87
Feb	24,831	22.09
Mar	17,199	15.29
Apr	9,818	8.73
May	15,701	13.97
Jun	21,402	19.04
Customer rating scale		
1	10,795	9.60
2	6,170	5.49
3	11,222	9.98
4	24,967	22.21
5	59,258	52.72

Note. The percentages were rounded up to two decimal point. Therefore, the percentage may not add to 100.0 because of rounding errors.

review came back smoothly to 14 % in May and almost 20 % in June.

4.2. Text-mining results

Four major features emerged from section 3.3.1 within the review comments including: 'service', 'food', 'place', and 'experience' (Table 3). Respectively, the feature of food indicated the tangible products that restaurants provided to their customers (e.g., "steak", "seafood", "wine", and "cocktails"). The feature of service (e.g., "waiter" and "delivery") described the whole process of interactions from the point when the customer got in touch with the restaurant until he or she had the product. The feature of place (e.g., "outside seating" and "neighborhood") described the geographical place and the physical display of a restaurant. The feature of experience (e.g., "value" and "satisfaction") represented a customer's general feelings or perceptions from the point when he or she entered the restaurant until they left.

4.3. Predicting Sentiment with deep learning models

Two machine learning and two deep learning algorithms were applied to test their abilities to classify customer sentiment mined from customer reviews using purely text-only messages, as shown in Table 4. Even though the models employed in the current study are some of the advanced recent techniques that were proven with high accuracy, the deep learning models outperformed machine learning models on test dataset. It is found that the Bidirectional LSTM achieved excellent performance, followed by Simple Embedding + Average Pooling. Two machine learning models performed worse than the state-of-the-art deep

Table 3
Summary of features mining in the Yelp reviews.

Features	Keywords
Service	staff, wait, shutdown, delivery, online ordering, accommodating, help, greeted, UberEats
Food	delicious, tasty, flavors, tender, options, sauce, menu, drink, fusion, appetizers
Place	street, station, close, parking, near, outside seating, location, convenience, easy, find
Experience	Value, disappoint, recommend, accepted, charge, atmosphere, clean, amazing, expectation, hygiene practices

Table 4
Summary of sentiment classification for four models.

	MSE	MAE	R^2	Fl
Random Forest	0.896	1.002	0.214	0.866
GBDT	0.883	0.798	0.338	0.889
Simple Embedding + Average Pooling	0.909	0.821	0.514	0.900
Bidirectional LSTM	0.917	0.913	0.601	0.920

learning models based on the four measures.

4.4. Sentiment analysis results

With the superiority of Bidirectional LSTM performance, the results of the sentiment score of four extracted factors between January and June 2020 are shown in Fig. 1. The sentiment scores ranged between -1 and 1, as -1 being the most negative and 1 being the most positive. Yelp users were overwhelmingly positive about their overall dining experience in the first half of 2020. However, there was a sharp decrease in sentiment scores in all four features in April. The sentiment scores for each feature started to bounce back in May.

4.5. Predicting review rating with deep learning models

Rating prediction performance on current Yelp datasets is presented in Table 5. In general, Simple Embedding + Average Pooling outperformed all other baseline models. The Bidirectional LSTM deep learning models also provided acceptable accuracy, which allowed even more significance with an average of about 10 % increase MSE and MAE compared with machine learning models. In general, the lower the MSE and MAE, the higher the performance of prediction. For both deep learning models, the values of MSE were 0.794 and 0.841, and MAE were 0.763 and 0.713.

Although the Bidirectional LSTM achieved an R^2 around 0.549, it was not perfect when compared to the Simple Embedding + Average Pooling (e.g., Socher et al., 2013). The findings are consistent with existing studies on deep learning models (Batmaz et al., 2018), which prove our results are fairly reliable. The predictive deep learning model is clearly very useful because it yields higher F1-measurements.

5. Discussion

This study empirically analyzed online restaurant reviews from Yelp in the era of the COVID-19 pandemic using traditional machine learning methods as well as deep learning methods. Based on the number of restaurant reviews posted on Yelp, an observable decline in March and a sharp decline in April were consistent with the timeline of how the COVID-19 outbreak evolved in the U.S.. Although the U.S. outbreak was officially declared a public health emergency on January 31, president
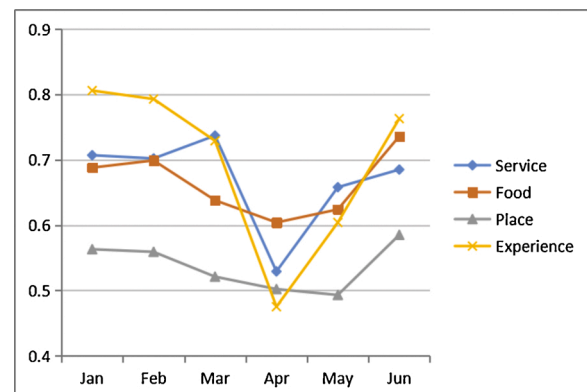

Fig. 1. Monthly sentiments of four features.

Table 5
Summary of review model accuracy.

Method	MSE	MAE	R^2	F1
Random Forest	1.029	0.823	0.298	0.890
GBDT	0.974	0.897	0.251	0.875
Simple Embedding + Average Pooling	0.794	0.763	0.607	0.911
Bidirectional LSTM	0.841	0.713	0.549	0.908

Trump declare a national emergency until March 13 (Whitehouse, 2020), which was two days after World Health Organization's (WHO) assessment that COVID-19 can be characterized as a pandemic. In April, it was reasonable that all four topics reached the lowest sentiment scores for the first half of the year since the pandemic reached its peak level (Kim, 2020), which made it even harder for restaurants to maintain their quality standards with limited staff and budget. In May, the number of online reviews began to bounce back since most of the states were gradually reopening businesses and customers were getting back to their pre-pandemic activity.

In terms of diners' sentiments towards restaurant, it was found that star ratings were heavily skewed toward the 4 and 5 stars. This study further extracted four main topics from restaurant reviews: service, food, place, and experience. Consistent with prior studies, food and service were two salient factors that influenced the customers' evaluation of a restaurant (e.g., Gan et al., 2017; Pantelidis, 2010; Zhang et al., 2014a, 2014b). With few exceptions (Bilgihan et al., 2018; Hyun, 2010), place is not frequently mentioned by prior studies as a dining experience attribute. When taking a look at the keywords falling under each topic, it was found that several barely mentioned keywords emerged during the COVID-19 pandemic. In terms of service, "shutdown", "delivery", "online ordering", and "UberEats" were frequently mentioned by customers. Amid the COVID-19 pandemic, a number of restaurants had to shut down if they had a positive COVID-19 case. Also, a lot of states limited restaurants to takeout and delivery only. Therefore, service quality was more likely to be associated with take-out and online food delivery services rather than dine-in services. It is interesting to find that customers evoked a positive emotion at a higher extent regarding service in March, with less positive emotions in the other three topics when compared to the previous two months. The potential reason might be that customers tended to take a restaurants' precarious position into account before evaluating the service quality of a restaurant. In terms of place, "outdoor seating" was extracted. This is another survival option for restaurants beyond offering take-out meals. This study proposed experience as a factor, which covers previously mentioned topics such as "price" (e.g., Hyun, 2010; Iglesias and Guillen, 2004), "atmosphere", and "environment" (e.g., Mattila, 2001; Zhang et al., 2014a, 2014b). "Hygiene practices" was associated with customers' dining experience, suggesting customers cared about whether the restaurant staff have good personal hygiene practices since it is critical in reducing the risk of spreading the disease. Furthermore, although restaurants reopened with capacity limits and social distancing rules in May, it was found that customers showed an overall positive attitude towards their dining experiences. Sentiments score of their attitudes towards food and place reached to the highest point in June, suggesting that restaurants are doing a good job maintaining consistency in their performance standards.

This study predicts sentiments and review rating using deep learning methods. It was found that overall deep learning models yielded better results when compared to machine learning algorithms. Since this study conducted an analysis on a large online review dataset (n = 112,412), the performance of deep learning algorithms was better than that of machine learning algorithms when the data volumes were large (Xin et al., 2018).

5.1. Theoretical implications

The study broadens the research on big data analytics in tourism and hospitality industry. From a methodological perspective, this study conducted a comparative analysis of machine learning methods and deep learning methods solving the sentiment analysis and review rating prediction task for restaurant online reviews. Most previous studies have either relied on qualitative methods such as narrative analysis, the AnswerTree method (Chang et al., 2011; Hwang and Zhao, 2010), or mixed methods by combining in-depth interview and survey (Cao et al., 2019) to identify the factors that are salient in evaluating customers' dining experience. These methods are subject to shortcomings such as costs and confirmation bias (Alaei et al., 2019). This study proposed two deep learning algorithms, deep learning average pooling and bidirectional LSTM, which could be utilized by future hospitality researchers to discover dynamics based on large sets of data without human involvement.

5.2. Practical implications

This study suggested that customers tend to focus on four features: service, food, place, and experience. Keyword lists for each feature were provided, which could help restaurateurs improve and expand on their current products and functionality. For example, customer perceptions of service quality are not only related to dine-in experience, but also online food delivery services. Therefore, restaurateurs should be cautious while using third-party delivery services as customers might blame the source restaurant for a late delivery, spilled food, or food delivered at a wrong temperature. Through sentiment analysis, it was observed that customers tend to be less satisfied about 'place; in the first six months of 2020, except for April. Although the location of the restaurant cannot be changed, the restaurateurs could increase customers' positive sentiments towards 'place' by providing plenty of parking to accommodate customers. Additionally, providing outdoor seating could also enable customers to easily find the restaurant and to eat out in the fresh air. During the COVID-19 pandemic, in addition to take out and delivery options, a lot of restaurants are only allowed to operate outdoor service in belief that fresh air can help defeat the coronavirus. Therefore, the unavailability of outdoor seating could negatively impact customers' emotional sentiments about 'place'. In addition, the four extracted features could be utilized by online review platform practitioners to improve their online rating system by asking customers to rate service, food, place, and experience separately, which could help prospective customers assess what to expect from a restaurant in an effective and efficient manner. Also, online review platform practitioners could make use of the keywords associated with each feature to provide customers guidelines on how to write a high-quality and useful online restaurant review.

To solve the discrepancy between text review contents and rating score, the online review platform practitioners could utilize deep learning average pooling techniques. A system could be developed to provide a warning to restaurant reviewers if there is a review–rating mismatch. In face of the COVID-19 crisis, as 92 % of restaurants reported using third-party delivery services in mid-March, a 27 % increase from the pre−COVID-19 period (Settembre, 2020), this method could also be utilized by third-party food delivery apps to detect online scam and spam reviews to ensure transparency and fair competition for the restaurant industry.

6. Conclusions, limitations, and future studies

This study proposed two deep learning algorithms to understand online restaurant reviews using the Yelp dataset. The Bidirectional LSTM algorithm was proven to be more effective in generating subtopics as well as sentiment prediction, while Simple Embedding + Average Pooling performs better in online review rating prediction tasks. This

study also found that the COVID-19 pandemic impacted restaurant rating and review trends.

This study has several limitations that need to be addressed. First, this study used reviews on Yelp that were posted during the first six months of 2020 to examine how customers' dining experience changes during the COVID-19 pandemic; however, Yelp does not ask customers to provide the actual visit date, which might lead to biased results. Second, this study only focused on data from restaurants located in five cities and one review platform. The robustness of the proposed deep learning model needs to be tested across different restaurant locations and across different review platforms. Third, although the deep learning technique provides researchers an effective way to make inferences about a large volume of complex data, it is often criticized for being a black box model with unknown and untraceable predictions (Buhrmester et al., 2019). Fourth, this study only included the Yelp dataset covering the first half of 2020, and future studies could use datasets from across longer time intervals to examine how the pattern of online restaurant review changes in the pre-, during-, and post-pandemic period.

Acknowledgement

References

Adi, Y., Kermany, E., Belinkov, Y., Lavi, O., Goldberg, Y., 2016. Fine-grained analysis of sentence embeddings using auxiliary prediction tasks. arXiv preprint arXiv: 1608.04207.

Aggarwal, C.C., Zhai, C. (Eds.), 2012. Mining Text Data. Springer Science & Business Media.

Ain, Q.T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., Rehman, A., 2017. Sentiment analysis using deep learning techniques: a review. Int. J. Adv. Comput. Sci. Appl. 8 (6), 424–433.

Alaei, A.R., Becken, S., Stantic, B., 2019. Sentiment analysis in tourism: capitalizing on big data. J. Travel. Res. 58 (2), 175–191.

Antonio, N., de Almeida, A.M., Nunes, L., Batista, F., Ribeiro, R., 2018. Hotel online reviews: creating a multi-source aggregated index. Int. J. Contemp. Hosp. Manage. 30 (12), 3574–3591.

Asghar, N., 2016. Yelp Dataset Challenge: Review Rating Prediction. Retrieved from. Cornell University. https://arxiv.org/pdf/1605.05362.pdf.

Bartik, A.W., Bertrand, M., Cullen, Z.B., Glaeser, E.L., Luca, M., Stanton, C.T., 2020. How Are Small Businesses Adjusting to covid-19? Early Evidence From a Survey(NBER Working Paper 26989). Retrieved from the National Bureau of Economic Research. National Bureau of Economic Research, Cambridge, MA. https://www.nber.or g/papers/w26989.pdf.

Bassig, M., 2019. The Impact of Restaurant Star Ratings on Customers. April Retrieved from. ReviewTrackers. https://www.reviewtrackers.com/blog/restaurant-star-rat ing/.

Bilgihan, A., Seo, S., Choi, J., 2018. Identifying restaurant satisfiers and dissatisfiers: Suggestions from online reviews. J. Hosp. Marke. Manage. 27 (5), 601–625.

Batmaz, Z., Yurekli, A., Bilge, A., Kaleli, C., 2018. A review on deep learning for recommender systems: challenges and remedies. Artif. Intell. Rev. 52 (1), 1–37.

Brown, J., Broderick, A.J., Lee, N., 2007. Word of mouth communication within online communities: conceptualizing the online social network. J. Interact. Mark. 21 (3), 2–20.

Buhrmester, V., Münch, D., Arens, M., 2019. Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: a Survey. arXiv preprint arXiv:1911.12116. Retrieved from. https://arxiv.org/abs/1911.12116.

Calheiros, A.C., Moro, S., Rita, P., 2017. Sentiment classification of consumer-generated online reviews using topic modeling. J. Hosp. Mark. Manage. 26 (7), 675–693.

Cao, Y., Li, X.R., DiPietro, R., So, K.K.F., 2019. The creation of memorable dining experiences: formative index construction. Int. J. Hosp. Manage. 82, 308–317.

Centeno, R., Hermoso, R., Fasli, M., 2015. On the inaccuracy of numerical ratings: dealing with biased opinions in social networks. Inf. Syst. Front. 17 (4), 809–825.

Chang, R.C., Kivela, J., Mak, A.H., 2011. Attributes that influence the evaluation of travel dining experience: when East meets West. Tour. Manag. 32 (2), 307–316.

Chatmeter, 2020. IMPORTANT: Google Disabling All Reviews & Responses Until Further Notice. April Retrieved from. https://www.chatmeter.com/blog/important-google-disabling-all-reviews-responses-until-further-notice/.

Chen, L.C., Lee, C.M., Chen, M.Y., 2019. Exploration of social media for sentiment analysis using deep learning. Soft Computing 1–11.

Christian, H., Agus, M.P., Suhartono, D., 2016. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). ComTech: Comput. Math. Eng. Appl. 7 (4), 285–294.

Croft, J., 2020. Yikes! Yelp Says 60% of Restaurant Covid-19 Closures Are Permanent. July Retrieved from. Cable News Network Business. https://edition.cnn.com/2020 /07/25/business/restaurants-reopen-coronavirus-shutdown-trnd/index.html.

De Choudhury, M., Gamon, M., Counts, S., Horvitz, E., 2013. Predicting depression via social media. Proceedings of the International AAAI Conference on Weblogs and Social Media (Icwsm 13). Retrieved from http://course.duruofei.com/wp-conten t/uploads/2015/05/Choudhury_Predicting-Depression-via-Social-Media_ICWSM13. pdf.

Dube, K., Nhamo, G., Chikodzi, D., 2020. COVID-19 cripples global restaurant and hospitality industry. Curr. Issues Tour. 1–4.

Feng, J., Yu, Y., Zhou, Z.H., 2018. Multi-layered gradient boosting decision trees. Advances in Neural Information Processing Systems 3551–3561.

Gan, Q., Ferns, B.H., Yu, Y., Jin, L., 2017. A text mining and multidimensional sentiment analysis of online restaurant reviews. J. Qual. Assur. Hosp. Tour. 18 (4), 465–492.

Ganu, G., Elhadad, N., Marian, A., 2009. Beyond the stars: improving rating predictions using review text content. Proceedings of 12th International Workshop on the Web and Databases (WebDB 2009) 1–6. June Retrieved from. http://citeseerx.ist.psu. edu/viewdoc/download?doi=10.1.1.150.140&rep=rep1&type=pdf.

Geierhos, M., Bäumer, F.S., Schulze, S., Stuß, V., 2015. "I grade what I get but write what I think." Inconsistency analysis in patients' reviews. Proceedings of European Conference on Information Systems 2015 (Paper 55). Retrieved from. https://aisel.ai snet.org/cgi/viewcontent.cgi?article=1054&context=ecis2015_cr.

Ha, J., Park, K., Park, J., 2016. Which restaurant should I choose? Herd behavior in the restaurant industry. J. Foodserv. Bus. Res. 19 (4), 396–412.

Hackeling, G., 2017. Mastering Machine Learning with scikit-learn. Packt Publishing Ltd., Birmingham, UK.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Boosting and additive trees. The Elements of Statistical Learning. Springer, New York, NY, pp. 337–387.

Hernández-Méndez, J., Muñoz-Leiva, F., Sánchez-Fernández, J., 2015. The influence of e-word-of-mouth on travel decision-making: consumer profiles. Curr. Issues Tour. 18 (11), 1001–1021.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hu, N., Koh, N.S., Reddy, S.K., 2014. Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. Decis. Support Syst. 57, 42–53.

Hwang, J., Zhao, J., 2010. Factors influencing customer satisfaction or dissatisfaction in the restaurant business using AnswerTree methodology. J. Qual. Assur. Hosp. Tour. 11 (2), 93–110.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.

Hyun, S.S., 2010. Predictors of relationship quality and loyalty in the Chain restaurant industry. Cornell Hotel Restaur. Adm. Q. 51 (2), 251–267.

Iglesias, M.P., Guillen, M.J.Y., 2004. Perceived quality and price: their impact on the satisfaction of restaurant customers. Int. J. Contemp. Hosp. Manage. 16 (6), 373–379.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems(NIPS 2017), pp. 3146–3154. Retrieved from. https://papers.nips.cc/paper/ 6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

Kim, A.R., Lee, S.W., Park, J., Kang, H.G., Seong, P.H., 2013. Correlation analysis between team communication characteristics and frequency of inappropriate communications. Annal. Nucl. Energy 58, 80–89.

Kim, A., 2020. After Falling for Months, Covid-19 Hospitalizations in the US Are Nearing April's Peak. July Retrieved from. CNN health. https://www.cnn. com/2020/07/22/health/coronavirus-hospitalization-scn-trnd/index.html#:~: text=At%20the%20peak%20of%20the,Adm.

Konchady, M., 2006. Text Mining Application Programming. Charles River Media, Inc.

Kragen, P., 2020. Bad Yelp Reviews During Pandemic Add Insult to Injury for Struggling Restaurants. August Retrieved from. Times Leader. https://www.timesleader.com/ wire/coronavirus-wire/797712/bad-yelp-reviews-during-pandemic-add-insult-to-in jury-for-struggling-restaurants.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lo, A.S., Yao, S.S., 2019. What makes hotel online reviews credible? An investigation of the roles of reviewer expertise, review rating consistency and review valence. Int. J. Contemp. Hosp. Manage. 31 (1), 41–60.

Luca, M., 2016. Reviews, Reputation, and Revenue: The Case of Yelp.cOm. Retrieved from. Harvard Business School NOM Unit Working Paper, pp. 12–016. https:// reviewinc.com/wp-content/uploads/2018/11/12-016_a7e4a5a2-03f9-490d-b093 -8f951238dba2.pdf.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E.C., Wetzels, M., Pfann, G., 2013. More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. J. Mark. 77 (1), 87–103.

Ma, E., Cheng, M., Hsiao, A., 2018a. Sentiment analysis–a review and agenda for future research in hospitality contexts. Int. J. Contemp. Hosp. Manage. 30 (11), 3287–3308.

Ma, Y., Xiang, Z., Du, Q., Fan, W., 2018b. Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning. Int. J. Hosp. Manag. 71, 120–131.

Martín, C.A., Torres, J.M., Aguilar, R.M., Diaz, S., 2018. Using deep learning to predict sentiments: case study in tourism. Complexity 2008, 7408431.

Martinez-Torres, M.R., Toral, S.L., 2019. A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tour. Manag. 75, 393–403.

Mattila, A.S., 2001. Emotional bonding and restaurant loyalty. Cornell Hotel Restaur. Adm. Q. 42 (6), 73–79.

McKinsey and Company, 2020. Lives and Livelihoods: Assessing the Near-Term Impact of Coronavirus on Workers. April Retrieved from. https://www.mckinsey.com/industries/public-sector/our-insights/lives-and-livelihoods-assessing-the-near-termimpact-of-covid-19-on-us-workers?.

Mouthami, K., Devi, K.N., Bhaskaran, V.M., 2013. Sentiment analysis and classification based on textual reviews. In: 2013 international conference on Information communication and embedded systems (ICICES). IEEE, pp. 271–276.

Mudambi, S.M., Schuff, D., Zhang, Z., 2014. Why aren't the stars aligned? An analysis of online review content and star ratings. In: Proceedings of 2014 47th Hawaii International Conference on System Sciences. Waikoloa, HI, USA, pp. 3139–3147. January Retrieved from. https://ieeexplore.ieee.org/document/6758991?denied=.

National Restaurant Association (2020). Reopening & Recovery - Guidance and information for restaurants looking to safely reopen. Retrieved from https://restaurant.org/manage-my-restaurant/business-operations/covid19/recovery.

Nie, R.X., Tian, Z.P., Wang, J.Q., Chin, K.S., 2020. Hotel selection driven by online textual reviews: applying a semantic partitioned sentiment dictionary and evidence theory. Int. J. Hosp. Manag. 88, 102495.

Oshiro, T.M., Perez, P.S., Baranauskas, J.A., 2012. How many trees in a random forest? In: Perner, P. (Ed.), Lecture Notes in Computer Science: Vol 7376. Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Springer, Berlin, Heidelberg, pp. 154–168.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222.

Pantelidis, I.S., 2010. Electronic meal experience: a content analysis of online restaurant comments. Cornell Hosp. Q. 51 (4), 483–491.

Pranckevičius, T., Marcinkevičius, V., 2017. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Balt. J. Mod. Comput. 5 (2), 221–232.

Qiu, J., Liu, C., Li, Y., Lin, Z., 2018. Leveraging sentiment analysis at the aspects level to predict ratings of reviews. Inf. Sci. 451, 295–309.

Quan, C., Ren, F., 2014. Unsupervised product feature extraction for feature-oriented opinion determination. Inf. Sci. 272, 16–28.

Rafay, A., Suleman, M., Alim, A., 2020. Robust review rating prediction model based on machine and deep learning: Yelp dataset. In: Proceedings of 2020 International Conference on Emerging Trends in Smart Technologies (ICETST). IEEE, pp. 8138–8143.

Rao, H., Shi, X., Rodrigue, A.K., Feng, J., Xia, Y., Elhoseny, M., et al., 2019. Feature selection based on artificial bee colony and gradient boosting decision tree. Appl. Soft Comput. 74, 634–642.

Redell, N., 2019. Shapley Decomposition of R-Squared in Machine Learning Models. *arXiv preprint arXiv:1908.09718*. Retrieved from. https://arxiv.org/abs/1908.09718.pdf.

Rubin, A., 2020. Writing Yelp Reviews during COVID-19. May Retrieved from. Yelp. https://blog.yelp.com/2020/05/writing-yelp-reviews-during-covid-19.

Sánchez-Franco, M.J., Navarro-García, A., Rondán-Cataluña, F.J., 2019. A naive Bayes strategy for classifying customer satisfaction: a study based on online reviews of hospitality services. J. Bus. Res. 101, 499–506.

Schuckert, M., Liu, X., Law, R., 2015. Hospitality and tourism online reviews: recent trends and future directions. J. Travel Tour. Mark. 32 (5), 608–621.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.

Shan, G., Zhang, D., Zhou, L., Suo, L., Lim, J., Shi, C., 2018. Inconsistency investigation between online review content and ratings. Proceedings of the 24th Americas Conference on Information Systems (AMCIS 18) 23.

Seo, S., Huang, J., Yang, H., Liu, Y., 2017. Representation learning of users and items for review rating prediction using attention-based convolutional neural network. In: Proceedings of 3rd International Workshop on Machine Learning Methods for Recommender Systems (MLRec 2017). Houston, TX, USA. Retrieved from. https://pdfs.semanticscholar.org/4946/89f4522619b887e515aea2b205490b0eb5cd.pdf.

Settembre, J., 2020. Restaurant Owners During Coronavirus Say Third-Party Delivery Saved Business and Workers. July Retrieved from. FoxBusiness. https://www.foxbusiness.com/lifestyle/restaurant-owners-say-third-party-delivery-saved-business-and-workers.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 conference on empirical methods in natural language processing 1631–1642. Retrieved from https://www.aclweb.org/anthology/D13-1170.pdf.

Sontag, E., 2020. Where Restaurants Have Reopened Across the U.S.. June Retrieved from Eater. https://www.eater.com/21264229/where-restaurants-reopened-across-the-u-s.

Tang, D., Qin, B., Liu, T., Yang, Y., 2015. User modeling with neural network for review rating prediction. In: Proceedings of Twenty-Fourth International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina, pp. 1340–1346. Retrieved from. https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/11051/10849.

Tsang, A., Prendergast, G., 2009. Is a "star" worth a thousand words?: the interplay between product-review texts and rating valences. Eur. J. Mark. 43 (11/12), 1269–1280.

Valdivia, A., Hrabova, E., Chaturvedi, I., Luzón, M.V., Troiano, L., Cambria, E., Herrera, F., 2019. Inconsistencies on TripAdvisor reviews: a unified index between users and Sentiment Analysis Methods. Neurocomputing 353, 3–16.

Van Calster, B., 2019. Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important). J. Clin. Epidemiol. 116, 137.

White, M., 2020. The Top 10 Largest U.S. Cities by Population. March 05 Retrieved from. https://www.moving.com/tips/the-top-10-largest-us-cities-by-population/.

Whitehouse, 2020. Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak. March Retrieved from. https://www.whitehouse.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/.

Wieting, J., Bansal, M., Gimpel, K., Livescu, K., 2015. Towards Universal Paraphrastic Sentence Embeddings. *arXiv preprint arXiv:1511.08198*. Retrieved from. https://arxiv.org/abs/1511.08198.

Xiang, Z., Du, Q., Ma, Y., Fan, W., 2017. A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour. Manage. 58, 51–65.

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., et al., 2018. Machine learning and deep learning methods for cybersecurity. IEEE Access 6, 35365–35381.

Xun, J., Guo, B., 2017. Twitter as customer's eWOM: an empirical study on their impact on firm financial performance. Internet Res. 27 (5), 1014–1038.

Yallop, A., Seraphin, H., 2020. Big data and analytics in tourism and hospitality: opportunities and risks. J. Tour. Futures 6 (3), 257–262.

Zhang, Z., Zhang, Z., Law, R., 2014a. Positive and negative word of mouth about restaurants: exploring the asymmetric impact of the performance of attributes. Asia Pacific J. Tour. Res. 19 (2), 162–180.

Zhang, Z., Zhang, Z., Law, R., 2014b. Relative importance and combined effects of attributes on customer satisfaction. Serv. Ind. J. 34 (6), 550–566.

Zhang, K.Z., Cheung, C.M., Lee, M.K., 2014c. Examining the moderating effect of inconsistent reviews and its gender differences on consumers' online shopping decision. Int. J. Inf. Manage. 34 (2), 89–98.

Zhang, K.Z., Zhao, S.J., Cheung, C.M., Lee, M.K., 2014d. Examining the influence of online reviews on consumers' decision-making: a heuristic–systematic model. Decis. Support Syst. 67, 78–89.

Zheng, T., Wu, F., Law, R., Qiu, Q., Wu, R., 2021. Identifying unreliable online hospitality reviews with biased user-given ratings: a deep learning forecasting approach. Int. J. Hosp. Manag. 92, 102658.