Framework for Affinity-Based Personalized Review Recommendation

Duy-Tan Nguyen[1], Yossiri Adulyasak[1], Jean-François Cordeau[1], Warut Khern-am-nuai[2]

Abstract

Online review platforms have proliferated due to recent technological advances and consumers' increased dependence on each other's opinions in purchasing decisions. Nevertheless, users typically face an enormous number of online reviews on the platform and suffer from information overload. Unlike existing studies which rely mostly on popularity, crowd-based evaluation, or filtering methods, we propose a comprehensive framework comprising exploratory, predictive, and prescriptive approaches for personalized review recommendation based on user-review affinity. Indeed, this study seeks to identify and recommend reviews to each user based on the propensity that the user will vote for, comment on, or re-read those reviews, whereby user login time increases, which in turn correlates positively with user affinity for the platform. We hypothesize a conceptual model, run predictive models, and conduct prescriptive analytics on the log data of a large restaurant review platform in Asia and find that reviewer-user similarity are among the significant explanatory variables, which is in line with the region's culture. Built on the result of the explanatory analysis, machine learning-based predictive models are then developed to predict the likelihood that each user will interact with each review in each restaurant. Our prescriptive results show that the resultant affinity-based ranking can significantly increase user engagement with the platform.

Keywords: review recommendation, user affinity, machine learning, partial least square structural equation modeling

[1] HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montreal, QC H3T 2A7, Canada, {duy-tan.nguyen, yossiri.adulyasak, jean-francois.cordeau}@hec.ca

[2] Desautels Faculty of Management, McGill University, 1001 Sherbrooke St W, Montreal, QC H3A 1G5, Canada, warut.khern-am-nuai@mcgill.ca

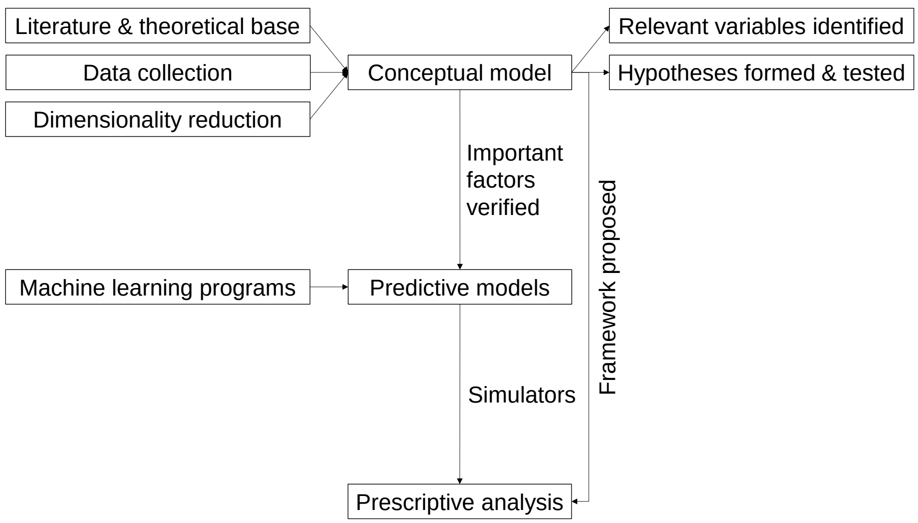
1. INTRODUCTION

In recent years, online review platforms have become one of the primary data sources for consumers (Huang et al. 2017). Prior studies have shown that many consumers rely heavily on online reviews in their decision-making process (e.g., Mudambi and Schuff 2010), resulting in strong empirical connections between online reviews and product sales (Chevalier and Mayzlin 2006, Sun 2012). For this reason, many platforms that aggregate online reviews have proliferated (Luca and Zervas 2016), especially for products that consumers cannot directly assess before consumption such as restaurants and hotels (Khern-am-nuai et al. 2018). These platforms have competed intensively to acquire and retain content contributors over the years (Qiao et al. 2020). However, such an increase in the amount of content also triggers an unforeseen issue for online review platforms. That is, users who face an enormous number of online reviews on the platform suffer from information overload (Gonzalez Camacho and Alves-Souza 2018), which generally leads to difficulty in filtering pertinent information (Zhou and Guo 2017). This issue has become increasingly important since recent studies have shown that such an issue may dampen the platforms' success (Chen et al. 2020). Several approaches have been utilized by online review platforms to mitigate this issue including the use of crowd-based content evaluation such as review helpfulness/unhelpfulness scores (Orlikowski and Scott 2014) and structured content filtering such as tags and badges (Rao et al. 2017). This research studies an alternative approach to help users overcome information overload in online reviews. Specifically, we propose a framework for personalized review recommendation system based on a theoretically driven design that can be readily implemented in online review platforms.

In the era of big data, the concept of personalized recommendation has been widely adopted by various services. Previous Information Systems (IS) literature has covered the design and use of personalized recommendation in multiple contexts including product advertising (Fleder and Hosanagar 2009, Hosanagar et al. 2014), news media (Prawesh and Padmanabhan 2014), and crowdsourcing contests (Mo et al. 2018). Nevertheless, interestingly, the use of personalized recommendation in online reviews is relatively rare in both research and practice. Meanwhile, existing works that specifically study personalized review recommendation are mostly limited to a certain aspect of online reviews such as review sentiment

(Zhang et al. 2018, Huang et al. 2020), consumer segment (Salehan et al. 2017), and review quality (Paul et al. 2017). Inspired by this gap in the literature and practice, this paper utilizes a unique dataset obtained through a collaboration with a large restaurant review platform in Asia to propose a comprehensive framework for personalized review recommendation based on user-review affinity, which can be broadly characterized as a user's positive attitude toward media content (Ji and Wayne Fu 2013). We specifically discuss and describe the quantitative measure of user-review affinity used in this work in Section 3.2.

Figure 1. Research design



Our research design consists of three phases: exploratory, predictive, and prescriptive (See Figure 1 for an overview). First, we extensively review prior online review studies to identify factors that potentially impact user-review affinity. Particularly, the exploratory phase of this study uses the partial-least square structural equation modeling or PLS-SEM (Henseler et al. 2016), which is commonly used for hierarchical models in big data analytics (Akter et al. 2017), to test a conceptual model that connects explanatory variables to user-review affinity. Following that, in our predictive phase, we develop a predictive model using machine learning algorithms and factors identified in the first phase as predictors to predict the likelihood that each user will interact with each review in each restaurant. Third, in our prescriptive phase, we rely on the results from the second phase to prescribe how reviews should be recommended to users on the platform and utilize counterfactual simulations, which have gained traction as a research methodology to validate prescriptive models in recent years (Sinai 2004, Xu et al. 2019), to measure the performance of our prescribed review recommendations.

The contributions of this work are threefold. First, we identify and verify important variables that are theoretically and empirically supported in determining relevant and useful reviews which in turn help increase user affinity for the platform. Second, we show that several machine learning models built on the verified factors can achieve comparable predictive performance at a lower computational cost vis-à-vis their high-dimensional counterparts. Finally, we demonstrate that arranging reviews in descending order of their predicted impact on user affinity rather than in chronological order is more effective in improving user affinity.

The rest of the paper is organized as follows. In Section 2, we review background literature that is related to this study. Then, we describe the data and our research context in Section 3. Section 4 discusses our exploratory model, followed by the predictive model in Section 5 and the prescriptive model in Section 6. Lastly, we conclude our study and discuss limitations and future research avenues in Section 7.

2. RELATED LITERATURE

In this section, we survey background literature that is related to our study. Particularly, we first review IS studies on recommender systems and then discuss prior works that specifically study personalized review recommendations.

2.1 Recommender Systems

There is an extensive body of literature on recommender systems (Mo et al. 2018), which have been strategically deployed by businesses to provide relevant recommendations to customers based on their purchasing histories and preferences (O'Mahony and Smyth 2010, Gorgoglione et al. 2019). According to Eirinaki et al. (2018), the most commonly used techniques include content-based, which analyzes a user's historical activities, and collaborative, which is based on other users with similar interests. In addition to the users' demographic profiles, germane details to build the model can come from the users' own reviews, search history (Bai et al. 2017), or their social networks (Li et al. 2017). Gonzalez Camacho and Alves-Souza (2018) find that social networks parlayed in collaborative algorithms are useful to make recommendations to new users or those with incomplete profiles, where preferences are not indicated, or to recommend new items to existing users, who might be interested in trying those products. In fact, to make recommendations for a new user, Son (2015) proposes a procedure which leverages similar users and similar items to their previously purchased products to predict ratings of a set of items for the user in question.

Nevertheless, such recommender systems must consider the accuracy-diversity dilemma since popular articles in peers' profiles may not perfectly match the user in question, which requires diversifying the algorithm into identifying or exploring items that might be better suited to the targeted customer's idiosyncrasies (Zhang et al. 2017). Moreover, Zhang et al. (2017) raise the caveat that recommender systems must attend to data recency in that user tastes and preferences evolve over time. Indeed, recent research has taken account of the evolution of both sellers and buyers, which may have emerged from their past interactions, to make recommendations (see Malgonde et al. 2020), but there is overall a lack of studies from business perspectives where user-centric and business-centric goals, e.g., satisfaction and profitability, must be considered (Gorgoglione et al. 2019). Given that recommender systems aim to help customers improve experiences and interact with businesses, where interactions include browsing, purchasing, and giving feedback (Gorgoglione et al. 2019), our paper focuses on predicting the probability that a user in question will interact more with a given review of a certain business via voting for, commenting on, or re-reading that review. This focus will be further justified and elaborated on in Section 3.

2.2 Personalized Review Recommendations

User-generated reviews which are often provided along with product recommendation have become an important source of information for customer decision-making and there has been ongoing research on personalized review recommendation (Mudambi and Schuff 2010, O'Mahony and Smyth 2010).

Wu (2017) finds that in determining review effectiveness for sales conversion, review popularity is as important as review helpfulness, which emphasizes the relevance to the user under analysis. Moreover, as the country, where the online review platform under analysis is based, scores high on collectivism (Hofstede 2001), implying a strong inclination toward conformity (Tsao et al. 2015), review popularity in collaborative-based recommendation systems can be relevant.

As user tastes and preferences evolve over time (Zhang et al. 2017), newly posted reviews are deemed more helpful by review readers (Hu et al. 2008, Filieri et al. 2015, Zhou and Guo 2017). Moreover, characteristics of the review itself and its reviewer are also found significant in the perceived helpfulness of the review and its impact on sales in several studies (Fang et al. 2016, Hu and Chen 2016, Hu et al. 2017), many of which are confirmed in Hong et al.'s (2017) meta-analysis of 47 pertinent publications. Moreover, Hong et al. (2017) confirm the moderating role of the platform host and the product category in determining the (perceived) helpfulness of the review. In effect, consumers consider reviews obtained from third-party platforms to be more reliable than those from seller-hosted platforms and experience products, whose quality evaluation is subjective and user-specific and thus hard to obtain via objective information search, entail consulting more online reviews, especially those whose reviewers have common interests and personalities with the reader/user (Hong et al. 2017).

Inasmuch as our data were collected from a third-party generated platform for online reviews on hospitality businesses, which mainly provide experience products, we can focus our model on attributes associated with reviews, reviewers, and users/readers.

3. DATA AND RESEARCH CONTEXT

3.1 Data Descriptions

We obtained our dataset through a collaboration with a large restaurant review platform in Asia. The platform currently has over three million users and over ten million reviews and photos for restaurants and other businesses (e.g., beauty salons and shopping malls) in approximately three hundred thousand locations in its home market.

Table 1. Descriptive statistics of the dataset

	Unique number	Number of interactions recorded				
		Mean	Standard deviation	5th percentile	Median	95th percentile
Users	216,556	8.47	10.24	2	4	59
Reviews	435,512	5.45	6.39	1	3	34
Reviewers	57,218	9.36	18.84	1	3	140
Businesses	76,703	23.78	38.94	2	8	234

In our dataset, which contains all reviews on the platform in 2017, there are 216,556 unique users, 435,512 reviews, 57,218 reviewers, and 76,703 businesses. In addition, there are 4,151,904 user-review interactions, including where reviews were read, voted for, or commented on by users. Descriptive statistics of interactions on the platform are provided in Table 1. We observe that users within the 5th–95th percentile had 8.47 interactions with reviews on the platform on average. However, the majority of them had no more than four interactions, suggesting that the data is heavily left-skewed. Meanwhile, from the review side, each review receives 5.45 interactions from platform users on average. Nevertheless, the majority of the reviews received no more than three interactions. In addition, at the aggregate level, reviewers on the platform had their reviews read/liked/commented on 9.36 times, on average, in 2017. In the meantime, businesses on the platform had their reviews read/liked/commented on 23.78 times, on average. The left skewness nature of the interactions is observed at the aggregate level too.

3.2 User-Review Affinity

The primary dependent variable of this study is user-review affinity. This variable is one of the main variables of interest to most online review platforms because it is strongly correlated with media-viewing time and frequency (Perse 1986, Ji and Wayne Fu 2013), which is directly connected to the platform's revenue and long-term sustainability. To operationalize the concept of user-review affinity, we consider three activities: vote, comment, and re-read as *indicators of increased user-review affinity* (hereafter referred to as *user-review affinity*). Specifically, user affinity for a review is equal to one if, within seven days after the initial read, the user either votes for a review, comments on a review, or re-reads a review, and zero otherwise. We select the seven-day threshold because the Ebbinghaus forgetting curve becomes relatively flat after such a period (Wixted and Ebbesen 1997, Li 2018).

From the users' perspective, these three activities strongly indicate their affinity toward reviews. For instance, helpfulness votes indicate that users find the corresponding review helpful for their decision-making (Tsai et al. 2020, Filieri et al. 2021). As a result, voting for a review explicitly indicates that users find useful information from the reviews provided, which, in turn, increases their affinity toward the review and the platform. In addition, prior literature has shown that helpful reviews tend to receive user comments (Malik and Hussain 2018). As such, prompting a reader to leave a comment can theoretically herald as review relevance and increased user affinity. Lastly, prior works also consider readership for review evaluation (Chua and Banerjee 2017). When users re-read the same review within seven days, they would visit the platform more and spend more time on the same content, which can also be considered increased affinity. Additionally, from the platform's perspective, these three activities are also used by our collaborating platform as a key performance measure because users tend to stay longer and access the platform more often when they vote, comment, or re-read reviews. These activities are also in line with the construct of user affinity in the literature (e.g., Sivasubramaniam and Chandrasekar 2019).

With the defined dependent variable of interest, the core idea of our model is to recommend reviews that are likely to attain high affinity to the users (i.e., reviews which users are likely to like, vote, or re-read). We next develop a framework to generate personalized review recommendations based on user-review affinity. We begin by exploring prior literature on factors that may affect user-review affinity, which we describe in the next section.

4. EXPLORATORY MODEL

In this section, we describe our exploratory model that is built to identify factors that could influence user-review affinity. In that regard, we draw on multiple prior works that show the influence of numerous variables on user affinity. Nevertheless, most works only consider the impact of these factors individually. The purpose of our exploratory model is to develop an integrated framework that verifies the effect of these factors when they are considered together. Results from this exploratory model will later be used to inform our downstream predictive and prescriptive models.

Here, we construct the problem at hand as a classification problem (i.e., the target variable captures whether a user would like/vote/re-read a review or not). In this regard, we follow the three fine-grained steps proposed by Mathias et al. (2013) to build a model for classification problems, including feature extraction, dimensionality reduction, and classification. User (she) and Reviewer (he) are denoted as the focal user and reviewer, respectively, in each datapoint.

4.1 Explanatory Variables and Dimensionality Reduction

In line with the literature (Hong et al. 2017, Malik and Hussain 2018, Liang et al. 2019, Hu and Yang 2021), the independent variables in our conceptual model belong to three groups, namely review features, reviewer characteristics, and product attributes. Based on an extensive survey of prior studies, we compile the list of variables used in this study as reported in Table 2. These variables individually influence user affinity according to prior literature.

Table 2. List of variables

	Definition/Operationalization/Feature	Prior studies
Review features	Review valence: the star ratings of the review (range between 1 and 5).	Purnawirawan et al. (2015), Quaschning et al. (2015)
	Review positivity: 1 if the rating is greater than 3, -1 if less than 3, 0 otherwise.	Sparks and Browning (2011), Filieri et al. (2021)
	Difference between review valence and business average rating.	Hu et al. (2008), Zhang et al. (2013), Fang et al. (2016)
	Difference between review valence and reviewer average rating.	
	Difference between review positivity and business average rating positivity.	
	Difference between review positivity and reviewer average rating positivity.	
	Review variance: The absolute difference between review rating and business average rating.	Quaschning et al. (2015), Xiang et al. (2017)
	Review helpfulness score: The (average) number of (helpfulness) votes that the review received.	Hu and Chen (2016), Zhou and Guo (2017), Wu (2017)
	Review age: The time difference between review posting and review reading.	Hu and Chen (2016), Gao et al. (2017), Hong et al. (2017)
	Review length: The number of words/characters in the review (measured by the platform in question)	Gao et al. (2017), Karimi and Wang (2017), Aghakhani et al. (2021)
	Review picture: The number of pictures in the review	Yang et al. (2017), Ma et al. (2018), Filieri et al. (2018)
Reviewer characteristics	Reviewer's total number of prior reviews.	Filieri et al. (2018), Zhou and Guo (2017)
	Reviewer's total number of reviews with quality flag	Filieri et al. (2019)
	Reviewer's total number of photos.	Filieri et al. (2018), Fang et al. (2016)
	Aka reviewer social network.	Hong et al. (2017), Zhou and Guo (2017)
	Reviewer's number of followers.	Hong et al. (2017), Yu et al. (2018), Aghakhani et al. (2021).
	Reviewer's number of followees.	
	Reviewer's number of followings (unique followers and followees).	
	User started following Reviewer recently (becoming friends within one day, one week, two weeks, one month or three months before).	Qian et al. (2014), Lee et al. (2015), Wang et al. (2018a), Liu et al. (2019)
	Reviewer started following User recently (becoming friends within one day, one week, two weeks, one month or three months before).	
	User's votes (likes and dislikes) for Reviewer's posts (within one day, one week, two weeks, one month or three months before).	Yu et al. (2020)
	Reviewer's votes (likes and dislikes) for User's posts (within one day, one week, two weeks, one month or three months before).	
	User's comments on Reviewer's posts before.	
	Reviewer's comments on User's posts before.	
	User-Reviewer common followers.	Xu et al. (2015)
	User-Reviewer common followees.	
	User-Reviewer indirect followships: Reviewer's followers are User's followees and vice versa.	
	Reviewer locality: If the reviewer is a local in the region of	Yang et al. (2017)

	Definition/Operationalization/Feature	Prior studies
	the reviewed business, reviewer locality is 1, 0 otherwise.	
Product attributes	Brand strength: The business average rating.	Ho-Dac et al. (2013), Blal and Sturman (2014), Tsao et al. (2019)

We next develop an integrated model that verify the influence of Table 2's variables together on user affinity. However, since there are 29 variables, we face two issues. First, as demonstrated in Lin et al. (2013), having that many exploratory variables with millions of observations would most likely result in an overfitted model. Second, having too many input variables can cause operational issues as well. For example, in the Random Forest Classifier model with M trees, n instances per decision tree, and $mtry$ features per tree, the algorithm complexity is $O(M \cdot mtry \cdot n \cdot \log n)$ (Wang et al. 2018c). While M and n are hyper-parameters to fine-tune in the classification step, the existing literature demonstrates that $mtry$ should equal $\sqrt{TotalFeatures}$ (Wang et al. 2018b). Taking both issues together, we proceed by performing dimensionality reduction to improve model identification and computational efficiency.

To ensure relevant variables are included while multicollinearity is avoided, we first conduct exploratory factor analysis (EFA). Eigenvalues and factor rotation (Yong and Pearce 2013) are used to select the high-order level factors that can capture most of the original variables not only for dimensionality reduction as in principal component analysis (Mason and Perreault 1991), but also for identification of latent features underlying certain sets of variables (Yong and Pearce 2013). We only retain factors whose Cronbach's alpha exceeds 0.7 (Dunn et al. 2014, Hair et al. 2019) and which comprise at least two items with loadings greater than 0.7 in absolute value. Table 3 illustrates the latent variables arising from our factor analysis, which are interpreted based on the literature. After EFA and PLS-SEM, we find ten composite scores that satisfy the convergent and discriminant validity criteria.

Table 3. Composite scores from EFA and PLS-SEM

Composite scores	Attributes	Loadings
Review valence frame* AVE = 0.83; CR = 0.97 α = 0.9606 Scale Corr. = 0.9988	Review valence ($\mathrm{rValence}$)	0.935
	Review positivity ($\mathbb{I}(\mathrm{rValence} > 3) - \mathbb{I}(\mathrm{rValence} < 3)$)	0.954
	Difference between review valence and business average rating	0.882
	Difference between review valence and reviewer average rating	0.883
	Difference between review positivity and business average rating positivity	0.915
	Difference between review positivity and reviewer average rating positivity	0.911
Reviewer expertise* AVE = 0.87; CR = 0.96 α = 0.9668 Scale Corr. = 0.9818	Log of Reviewer's total number of prior reviews	0.990
	Log of Reviewer's total number of reviews with quality flag	0.993
	Log of Reviewer's total number of photos	0.963
	Log of Reviewer's total number of followers	0.769

Reviewer-user similarity* AVE = 0.77; CR = 0.98 α = 0.9727 Scale Corr. = 0.9511	Log of User's likes for Reviewer's posts before	0.951
	Log of User's votes for Reviewer's posts before	0.951
	Log of User's comments on Reviewer's posts before	0.834
	Log of Reviewer's likes for User's posts before	0.950
	Log of Reviewer's votes for User's posts before	0.950
	Log of Reviewer's comments on User's posts before	0.816
	User's recent votes (likes) for Reviewer's posts	0.875
	Reviewer's recent votes (likes) for User's posts	0.863
	Log of User-Reviewer common followers	0.843
	Log of User-Reviewer common followees	0.782
	Log of User-Reviewer indirect followship 1	0.840
	Log of User-Reviewer indirect followship 2	0.864
Review quality** AVE=CR=α=Scale Corr=1	Log of review's average number of votes received	1.000
	Log of review's average number of likes received	1.000
Review votes (likes)** AVE=CR=α=Scale Corr=1	Log of review's number of votes received	1.000
	Log of review's number of likes received	1.000
User following Reviewer recently** AVE = 0.83; CR = 0.95 α=0.93; Scale Corr.=0.98	User started following Reviewer within 1 day before	0.827
	User started following Reviewer within 7 days before	0.941
	User started following Reviewer within 14 days before	0.959
	User started following Reviewer within 30 days before	0.913
Reviewer following User recently** AVE = 0.85; CR = 0.94 α=0.91; Scale Corr.=0.95	Reviewer started following User within 7 days before	0.888
	Reviewer started following User within 14 days before	0.953
	Reviewer started following User within 30 days before	0.916
Social connectedness** AVE = 0.92; CR = 0.96 α=0.93; Scale Corr.=1.00	Log of Reviewer number of followees	0.930
	Log of Reviewer followings	0.990
User's recent votes (likes) for Reviewer's posts*** α = 0.9926 Scale Corr. = 0.9983	Log of User's likes for Reviewer's posts within 7 days	0.953
	Log of User's likes for Reviewer's posts within 14 days	0.987
	Log of User's likes for Reviewer's posts within 30 days	0.988
	Log of User's likes for Reviewer's posts within 90 days	0.960
	Log of User's votes for Reviewer's posts within 7 days	0.953
	Log of User's votes for Reviewer's posts within 14 days	0.987
	Log of User's votes for Reviewer's posts within 30 days	0.988
	Log of User's votes for Reviewer's posts within 90 days	0.960
Reviewer's recent votes (likes) for User's posts*** α = 0.9910 Scale Corr. = 0.9982	Log of Reviewer's likes for User's posts within 7 days	0.939
	Log of Reviewer's likes for User's posts within 14 days	0.983
	Log of Reviewer's likes for User's posts within 30 days	0.985
	Log of Reviewer's likes for User's posts within 90 days	0.954
	Log of Reviewer's votes for User's posts within 7 days	0.939
	Log of Reviewer's votes for User's posts within 14 days	0.983
	Log of Reviewer's votes for User's posts within 30 days	0.985
	Log of Reviewer's votes for User's posts within 90 days	0.954

Note: α = Cronbach's alpha
* *Latent variable*: aggregate variable created in PLS-SEM that are also supported in SEM.
** *Composite score*: aggregate variable created in PLS-SEM (Hair et al. 2020).
*** *Item parceling*: aggregating items into a parcel which is used as an indicator in SEM (Hall et al. 1999).

Review valence, review positivity, and its difference from business/reviewer average rating (positivity) are highly correlated and altogether reflective of whether a review is in favor of the reviewed item/business vis-à-vis other reviews, which we name *review valence frame*. The higher rating a reviewer gives, the more likely that rating outstrips the business/reviewer average rating, which implies his higher favor toward the business compared to an average reviewer and vice versa.

We can observe that the total number *of* prior reviews and the total number of prior reviews with quality flag (i.e., reviews that receive multiple helpfulness votes), which are visible in the system, can serve as cues for readers about the reviewer's expertise and the likely helpfulness of the focal review. More specifically, Reviewer's expertise relates to his number of previous reviews and helpfulness votes (Zhou and Guo 2017, Filieri et al. 2019). Given that reviews posted with photos are likely deemed helpful (Fang et al. 2016, Filieri et al. 2018), reviews with many reviews with quality flag might have posted many pictures. So, the number of his posted photos can indicate his expertise. In line with Yu et al. (2018), reviewer number of followers correlates strongly with these reviewer expertise attributes in our data. Hence, the latent attribute underlying these variables can be interpreted as *reviewer expertise*.

Next, Neirotti et al. (2016) find that when a user and reviewer share similar interests or know each other, they tend to trust the review and likely like, comment on, or re-read it. On the partner platform, these attributes can be captured by prior interactions such as comments and likes for previous reviews between a user and reviewer, whose latent variable is supposedly interpreted as reviewer-user interactions. Yet, we also find that common followship features, which are measured by the common followers and followees of a reviewer and user, are also captured by the same latent variable as reviewer-user interactions attributes. Since these common followship measures are also utilized to reflect the similarity between a focal user and a potential followee in the followee recommendation literature (Xu et al. 2015), the latent variable that captures both reviewer-user interactions and common followship can be interpreted as *user-reviewer similarity*.

Next, we run PLS-SEM and perform confirmatory composite analysis (Hair et al. 2020) to assess the EFA results. We choose PLS-SEM, whose add-in package for STATA was developed by Venturini and Mehmetoglu (2019), because this method allows testing theoretical models for predictive purposes, relaxing normality assumptions, and utilizing the latent scores for subsequent analyses (Hair et al. 2019). We conduct convergent validity analysis (Sethi and King 1994) and discriminant validity analysis (Fornell and Larcker 1981), which are commonly adopted in the literature (e.g., Henseler et al. 2016, Hair et al. 2019) to substantiate the scale validity. As observed in Table 3, all standardized path coefficients λ_{ij} for component j in latent variable/composite score i are of acceptable magnitude and statistically significant, implying good convergent validity (Sethi and King 1994, Hair et al. 2019). With respect to the discriminant validity, Table 3 shows that all factors have good composite reliability, which is above the 0.7 threshold (Fornell and Larcker 1981, Hair et al. 2020). Meanwhile, the Average Variance Extracted of each factor exceeds the 0.5 threshold and its squared correlations with other factors (see Table 4), which is another indicator of good discriminant validity (Henseler et al. 2016, Hair et al. 2020). By only considering factors in EFA with Cronbach's alpha greater than 0.7, which is regarded as a lower bound to internal consistency of a factor (Sijtsma 2009, Henseler et al. 2016), we believe that the factors reported here are properly measured by their components, whose contents are relevant to the latent variables we target. We also ran covariance-based SEM, and the results were robust for the three latent variables in Table 3.

Table 4. Correlations between explanatory variables in the structural part of PLS-SEM

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1.00																
–.43	1.00															
–.01	–.05	1.00														
0.02	–.02	0.32	1.00													
0.01	0.02	0.17	0.47	1.00												
0.08	–.06	0.24	0.44	0.48	1.00											
0.05	0.09	–.29	0.31	0.16	0.03	1.00										
–.07	–.04	0.22	0.52	0.46	0.44	0.09	1.00									
–.02	–.03	0.48	0.13	0.09	0.11	–.13	0.15	1.00								
0.00	–.01	0.07	0.01	0.01	0.02	–.03	0.01	0.07	1.00							
–.01	–.01	0.09	0.02	0.01	0.02	–.03	0.03	0.22	0.01	1.00						
0.00	–.01	0.09	0.00	0.00	0.01	–.04	0.00	0.11	0.21	0.02	1.00					
–.01	0.00	0.09	0.03	0.02	0.02	–.02	0.03	0.21	0.01	0.06	0.01	1.00				
–.03	–.06	0.27	0.66	0.36	0.36	0.14	0.61	0.17	0.02	0.03	0.02	0.03	1.00			
0.00	–.01	–.01	–.02	–.02	–.02	–.02	–.03	0.01	0.00	0.00	0.00	0.01	–.02	1.00		
0.02	0.03	–.07	–.08	–.04	–.04	0.02	–.09	–.07	0.00	–.02	–.01	–.02	–.06	–.07	1.00	
0.28	–.11	0.02	–.01	–.02	0.04	0.02	–.12	–.01	0.01	0.00	0.00	0.00	–.05	0.01	0.02	1.00

Note: Values less than 1.00 on the diagonal are the Average Variance Extracted of the corresponding composite score or latent variable. (1) Review valence frame; (2) Review variance; (3) Review quality; (4) Review votes (likes); (5) Review length; (6) Review picture; (7) Review age; (8) Reviewer expertise; (9) Reviewer-user similarity; (10) User following Reviewer recently; (11) User dislikes for Reviewer; (12) Reviewer following User recently; (13) Reviewer dislikes for User; (14) Reviewer social connectedness; (15) Reviewer locality; (16) User-Reviewer common locality; (17) Brand strength.

Overall, the tests above corroborate the validity of the factors arising from our factor analysis. With an original set of variables, we develop eight composite scores and remove variables whose contents are related to the composite scores created. The eight composite scores include *review valence frame*, *review quality*, *review votes/likes*, *reviewer expertise*, *reviewer-user similarity*, *reviewer social connectedness*, *reviewer following user recently*, and *user following reviewer recently*. Moreover, there are nine variables, namely *review variance*, *review age*, *review length*, *review picture*, *reviewer dislikes for user*, *user dislikes for reviewer*, *reviewer locality*, *reviewer-user common locality*, and *brand strength* that are directly measured by one single feature and are not captured by the eight composite scores in our factor analysis.

Finally, we obtain a new model with 17 variables. We report the correlation of these variables in Table 4.

To combine the variables captured by a common factor, we test the correlation between their unweighted and weighted average and find a strong correlation of at least 0.95 in all cases (Scale Corr. in Table 3). Hence, we can proceed with the unweighted average to build our predictive model, which is then called '*model with unweighted scales*' because it is convenient to generate and repeat (see the discussion of Bobko et al. 2007). We also perform robustness checks by comparing the performance of this model and the one '*with weighted scores*' where the weight vector for a composite score's items is computed by the PLS-SEM algorithm (see Venturini and Mehmetoglu 2019). The results of this alternative model are qualitatively similar to those of our main model.

4.2 Hypothesis Development and Testing

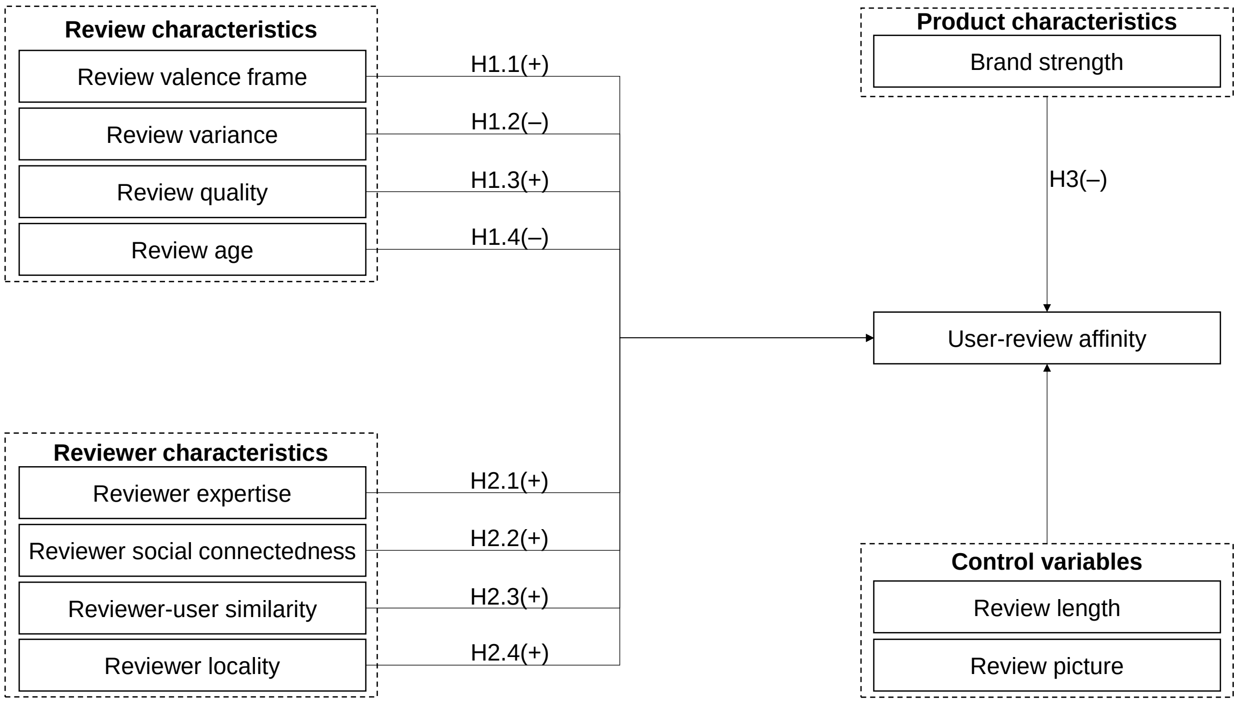
Now that we obtain the set of independent variables of interest, we next formally develop hypotheses and test whether they affect user affinity statistically. The conceptual model hypothesized for the three groups of variables identified in section 4.1, i.e., review attributes, reviewer characteristics, and product features, is illustrated in Figure 2. Sections 4.2.1–4.2.3 provide detailed elaboration on each hypothesis and section 4.2.4 presents the results of the hypothesis testing.

4.2.1 Review Characteristics

We begin by examining the influence of review characteristics on user-review affinity. The first characteristic of our interest is the review valence frame (i.e., the polarity of the review), which has been widely studied in the literature. More particularly, Purnawirawan et al. (2015) show in their meta-analysis that that review valence has significant impact on user-review affinity (in terms of review helpfulness votes) for experience goods and unfamiliar brands. In the same vein, Quaschning et al. (2015) find in their field and experimental data that the valence of a review significantly impacts its helpfulness when it is in accord with other reviews. Lee et al. (2017) expand on the effect of review valence on review helpfulness and show that reviews with negative valence are usually perceived as more helpful compared with reviews with positive valence. However, their helpfulness decreases when the negative emotions in the reviews are intense. Following these prior findings, we hypothesize:

H1.1. Review valence frame increases user-review affinity.

Figure 2. Conceptual Framework



Next, we investigate the influence of the consistency of review valence (i.e., the variance of the reviews) on user-review affinity. Indeed, Yelp's strong review helpfulness is ascribable to the high variance in its review sentiment (Xiang et al. 2017). Meanwhile, signaling theory suggests that if a review's valence is widely dispersed from the average rating, that inconsistency may signal the reviewer's idiosyncrasy/heterogeneity (Quaschning et al. 2015). In this regard, Gao et al. (2017) show that users are more likely to cast a helpfulness vote when there is consistency between the focal review's valence. Based on the aforementioned empirical evidence, we hypothesize that high review variance negatively affects review helpfulness as perceived by users. As such:

H1.2. Review variance decreases user-review affinity.

Many scholars, e.g. Hu et al. (2017), Lu et al. (2018), and Liang et al. (2019), find a direct and significant influence of review quality on review helpfulness, whereas the study of Lee et al. (2018) shows that review quality has poor predictive performance for review helpfulness. We operationalize review quality by the number of votes a review had already got before being read by User, which is supported in the literature (see Yang et al. 2017). On the platform in question, the number of review votes include the number of likes and the number of dislikes, so we incorporate all those figures into the model. Our dimensionality reduction shows that the total number of review votes and likes are captured by one composite scale, so are the average number of review votes and likes, which respectively equal the total number of review votes and likes divided by the time lapse in days since review post time. Given that old reviews have more time to accumulate votes, we use the average number of votes and likes to proxy review quality and to penalize less recent reviews as Hu et al. (2017), Lee et al. (2018), and Tsai et al. (2020) did.

In line with signaling theory, prior votes, likes, and dislikes are visible clues for readers about review quality. As the network effect and collectivism dictate, User is likely to conform with the majority and interact with such reviews by either voting for, commenting on, or re-reading it within seven days. As a result, we hypothesize that review quality increases user affinity for the platform.

H1.3. Review quality increases user-review affinity.

Another commonly discussed review feature is review age, which is measured in days elapsed since review post time (Hu and Chen 2016, Hong et al. 2017, Hu and Yang 2021) and can be rescaled by log-transformation (Gao et al. 2017, Aghakhani et al. 2021). While Hu and Chen (2016), Gao et al. (2017), and Hong et al. (2017) find that review age raises the perceived review helpfulness, the results of Wu (2017) are mixed, depending on the product type, but the aggregate effect (pooled across product types) is negative. Meanwhile, the study of Yang et al. (2019) illustrates a negative influence, which means that older reviews are considered less helpful. As the businesses reviewed on the platform in question provide hedonic and experience products, our hypothesis is in favor of the findings of Wu (2017) and Yang et al. (2019), where review age decreases the relevance/helpfulness of the reviews for such items. Given that User is less likely to cast a helpfulness vote or spend time commenting on or re-reading a less relevant/helpful review, we hypothesize that

H1.4. Review age decreases user-review affinity.

Our review platform arranges reviews in ascending order of review post time lapse (review position rank). We adopt that measure as a proxy for review age. Using that feature in our model also helps us with counterfactual analysis given that how reviews are arranged in the system can be manipulated.

In a separate vein of research, Quaschning et al. (2015), Fang et al. (2016), Gao et al. (2017), Wu (2017), and Hu and Yang (2021) illustrate that review length positively affects review helpfulness. Meanwhile, Zhou and Guo (2017) show a marginally significant moderating impact of review length, whereas Karimi and Wang (2017) and Zhang and Lin (2018) find a negative effect. Obviously, lengthy reviews are less likely to be perused and thus less likely to be assessed. The review length which is readily measured on the platform is log-transformed in our model as in the papers of Gao et al. (2017), Karimi and Wang (2017), and Aghakhani et al. (2021). We include this variable as a control variable and formulate no hypothesis for it. Nonetheless, by including this variable as a control variable, we can exclude the possibility that some reviews were re-read because they were lengthy.

Another control variable whose mean, median, and standard deviation are similar between the two groups of reviews in our database is *review picture*. According to Zhou and Guo (2017), Yang et al. (2017), and Ma et al. (2018), when a review is accompanied by at least one photo related to the reviewed item/business, users are more likely to perceive that review to be helpful. Filieri et al. (2018) also find that the perceived helpfulness of extreme reviews increases when they are long and posted with pictures since photos are deemed more convincing than words (Fang et al. 2016).

4.2.2 Reviewer Characteristics

According to Filieri et al. (2019) and Quaschning et al. (2015), reviews whose valence is inconsistent with the majority of other reviews are still perceived as helpful if the reviewers are considered expert. This signals that reviewer expertise is an important variable in predicting review helpfulness. Indeed, Hu and Yang's meta-analysis (2021) shows that the impact of reviewer expertise on review helpfulness is significant and positive yet declining over time. Zhou and Guo (2017) also find that reviewer expertise positively affects review helpfulness. In addition, reviewer expertise attenuates the effect of the number of prior reviews on the perceived helpfulness of the focal review (Zhou and Guo 2017). Similar results can be found in the papers of Hong et al. (2017), Lee et al. (2017), and Yang et al. (2019). Reviewer expertise is a multifaceted concept, which has been operationalized differently in the literature, and the effect of each reviewer expertise attribute on review helpfulness is mixed. Siering et al. (2018) and Yang et al. (2019), for instance, measure reviewer expertise by the reviewer rank computed on Amazon.com, whereas Filieri et al. (2019) operationalize that variable by the number of helpfulness votes the focal reviewer received. In Hong et al.'s meta-analysis (2017), reviewer expertise as operationalized by *expert title/label* has consistently positive impact on review helpfulness while the result for the total number of reviews posted as a reviewer expertise attribute is inconsistent. As measured by the number of helpfulness votes received, reviewer expertise positively impacts review helpfulness, whereas reviewer rank (reputation) has negative impact in Lee et al.'s study (2017) because their variable of interest is *emotional intensity* in negatively-valenced reviews, which may lower information diagnosticity perceived by readers. Filieri et al. (2018) report similar results and find that the number of reviews posted as a reviewer expertise element is statistically insignificant. Aghakhani et al. (2021) and Liang et al. (2019) even find a negative impact of that variable on review helpfulness. Zhou and Guo (2017) combine both the number of reviews posted and elite membership to proxy reviewer expertise and report an aggregate positive impact on review helpfulness.

The review platform under analysis has an indicator of reviewer rank, but we cannot backtrack its value to the time each review was read, so we do not include them in the model. In our work, reviewer expertise is measured by reviewer prior total votes, likes, photos, and followers as supported in the dimensionality reduction results. Overall, as reviewer expertise increases review helpfulness, User is more likely to cast a helpfulness vote, which then indicates her increased affinity for the platform. So, we hypothesize

H2.1. Reviewer expertise increases user-review affinity.

As studied by Filieri et al. (2019) and Quaschning et al. (2015), the possible interaction between review variance and reviewer expertise should be considered. We compute this interaction term by multiplying review variance (single item) by each standardized component of reviewer expertise (see Chin et al. 2003). This interaction term between review variance and reviewer expertise is the 18th variable in our model.

According to Zhou and Guo (2017), social connectedness in addition to reviewer expertise also has direct and moderating impacts on review helpfulness. Social connectedness is defined as the relationships a user has with other users on the platform and is operationalized as the number of friends on Yelp in Zhou and Guo's study (2017). Zhang and Lin (2018) consider both the number of friends and followers (fans) on Yelp in operationalizing reviewer social networks. On the review platform we study, this concept can be computed by the *number of followers* and the *number of followees*. Hong et al. (2017) find that the *number of followers* and the *number of followees* have consistently positive impact on review helpfulness. Aghakhani et al. (2021) log-transform these figures in their model, but the results are insignificant. Yu et al. (2018) consider these two indices in computing a user's expertise in a field. Let $fe(u_i)$ denote the set of users followed by user i on the platform and $fr(u_i)$ denote the set of users following user i on the platform. In our paper, *reviewer followings* $= fe(u_i) \cup fr(u_i)$, and reviewer social connectedness is measured by reviewer's number of followees and followings (see section 4.1). As can be inferred from the network effect, reviewers who have many followings are likely to write high-quality and valuable reviews, which can prompt readers to cast helpfulness votes, leave comments, or re-read, so our hypothesis is

H2.2. Reviewer social connectedness increases user-review affinity.

Other research streams have shown that users tend to follow their friends when rating an item or business (Lee et al. 2015, Wang et al. 2018a) or trying a new product/service, which is leveraged for personalized recommendation (Qian et al. 2014, Liu et al. 2019). As discussed by Neirotti et al. (2016), users assign greater weight to reviews written by friends in their network. Therefore, a follower-followee relationship or frequent interactions, i.e., votes for or comments on reviews, between User and Reviewer can signal that his review is more likely to be perceived as helpful by her.

Even if the user and the reviewer have not established a follower-followee relationship, we can identify potential followees based on the followee recommendation scholarship. As the followees recommended are likely to share common interests with the user (Armentano et al. 2013, Li et al. 2016), reviews posted by the former can be relevant and helpful to the latter. Since the candidate followees are not yet in the user's network, suggesting their reviews to the user can boost the diversity of review recommendation. To identify relevant followees, we can compute the similarity between the focal user (user i) and another user (user j), as adapted from Xu et al. (2015), by: $sameFe = \left|fe(u_i) \cap fe(u_j)\right|$; $sameFr = \left|fr(u_i) \cap fr(u_j)\right|$;

$indirectFollowship1 = \left|fe(u_i) \cap fr\left(u_j\right)\right|$; $indirectFollowship2 = \left|fr(u_i) \cap fe\left(u_j\right)\right|$.

By computing these indicators from the followee recommendation studies, we can find the common followship level between User and Reviewer and thereby predict review relevance. Both the network effect and collectivism can help justify this selection as User is apt to find shared values with her followers or followees, who then share commonality with Reviewer. This can be explained by the high collectivism in Southeast Asia, where members tend to find shared values within their group (Hofstede 2001); thus, a review written by a friend can be deemed helpful or relevant.

As indicated in our dimensionality reduction results, we conceptualize this variable as *reviewer-user similarity*. Given a high level of *reviewer-user similarity*, User may well find shared values directly via prior interactions or indirectly through common followships with Reviewer, so she is more likely to vote for, comment on, or re-read his review. Incorporating this variable in our model can help account for possible autocorrelation, where User would continue to vote for and comment on Reviewer's reviews as she did in prior observations. Further, frequent interactions between users of similar interest clearly boosts user emotional attachment to the platform. Therefore, we hypothesize

H2.3. Reviewer-user similarity increase user-review affinity.

Yang et al. (2017) find that reviews written by local reviewers are perceived as more helpful. Explanations may include Reviewer's hands-on experience in the region and in using the product of the business reviewed, which implies that his review is more credible. Hence, we hypothesize that reviews posted by reviewers from the same neighborhood are likely perceived as helpful. Additionally, as reviewer region ID is observable on the platform, User can check if Reviewer is from her locality. With high collectivism in Southeast Asia, where members tend to find shared values within their group (Hofstede 2001), we also hypothesize that if User and Reviewer have the same region ID, his review is more likely to be relevant to her, prompting her to cast a vote, leave a comment, or re-read. In other words,

H2.4. Reviewer locality increases user-review affinity.

4.2.3 Product Characteristics

As can be seen from our discussion in Section 3.1.1, *product type* plays the moderating role in the models of Wu (2017) and Mudambi and Schuff (2010). Nonetheless, as all the businesses reviewed on the platform deliver hedonic or experience products in the field of hospitality, product type cannot differentiate review helpfulness in our model and is thus not considered. Still, we can see in the cited literature some other less commonly controlled yet relevant features such as brand similarity (Purnawirawan et al. 2015), total reviews received (Lee and Choeh 2016, Filieri et al. 2021), product quality, awareness and popularity (Zhang and Lin 2018), hotel features (Liang et al. 2019, Filieri et al. 2021) and average rating (Filieri et al. 2021). We can see that these variables appertain to *brand strength*. According to Sridhar and Srinivasan's empirical findings (2012), customers' ratings are affected by peer pressure. Particularly, reviewers may give a higher rating than their actual product experience if prior ratings are positive (Sridhar and Srinivasan 2012). As a strong brand usually has good cumulative ratings, new customers may tend to follow that norm when rating. Another explanation relates to Tsao et al.'s findings (2019) that the impact of negative reviews on sales are stronger than that of positive ones for strong brands and that the management is recommended to address such negative reviews. As a result, the overall ratings are higher for stronger brands. Therefore, we operationalize brand strength by business average rating. There are mixed results for this variable in the literature. Purnawirawan et al. (2015) find that reviews for unfamiliar brands are deemed more helpful, whereas brand strength indices studied by Lee and Choeh (2016), Zhang and Lin (2018), and Filieri et al. (2021) positively affect or moderate review helpfulness, especially for experience goods, which are pertinent to the review platform under analysis in our study. The network effect and collectivism may predict that patronizing strong brands implies compatibility with peers or the majority, which is valued by User, so reading reviews or even accessing a review platform for information is less helpful. Therefore, we hypothesize

H3. Brand strength decreases user-review affinity.

We use the first six months of the data to test the conceptual model (Figure 2). The last six months' data will be used as an out-sample data to test the model's generalization to unseen instances. In this paper, except for binary and ordinal variables (e.g., rating), continuous variables are log-transformed and normalized.

4.2.4 Hypothesis Testing

Table 5 presents the tested results for the conceptual model, based on which important features are inputted into predictive models and prescriptive analytics. We compute the variance inflation factor (VIF) index to check if multicollinearity still persists after dimensionality reduction. Since the VIF is less than 10, the structure of our model is supported (Marquaridt 1970).

Table 5. PLS-SEM Path analysis

Number of observations	1813477	Absolute GOF	0.19552
Average R-squared	0.04537	Relative GOF	0.91812
Average communality	0.87763	Average redundancy	0.04537

Dependent variable = User-review affinity

Variable	Coefficient	$P > \lvert z \rvert$	VIF
(1) Review valence frame	0.0220	0.000	1.359
(2) Review variance	–0.0019	0.020	1.290
(3) Review quality	0.0724	0.000	1.714
(4) Review votes (likes)	–0.0110	0.000	2.474
(5) Review length	–0.0279	0.000	1.542
(6) Review picture	0.0006	0.531	1.508
(7) Review age	–0.0037	0.000	1.401
(8) Reviewer expertise	–0.0721	0.000	1.926
(9) Reviewer-User Similarity	0.1614	0.000	1.433
(10) User following Reviewer recently	0.0186	0.000	1.052
(11) User dislikes for Reviewer	–0.0079	0.000	1.049
(12) Reviewer following User recently	0.0141	0.000	1.060
(13) Reviewer dislikes for User	–0.0057	0.000	1.047
(14) Reviewer social connectedness	0.0340	0.000	2.193
(15) Reviewer locality	0.0026	0.000	1.007
(16) User-Reviewer common locality	0.0191	0.000	1.021
(17) Brand strength ($\mathrm{busAvgRating}$)	–0.0029	0.000	1.110
(18) Review variance × Reviewer expertise	0.0117	0.000	1.030

Turning first to the group of hypotheses on review features, Table 5 shows that review valence frame has a statistically significant and positive impact on user-review affinity, which means H1.1 is supported. In other words, positive reviews are more likely to improve user affinity to the platform.

As expected, review variance negatively impacts user-review affinity and H1.2 is supported. Albeit hedonic products imply heterogenous preferences (Yang et al. 2017), the network effect and conformity pressure prevail in Southeast Asia’s collectivism, resulting in users’ less favor for extreme reviews.

Prior studies used either review total votes/likes or review average votes/likes per day to measure review quality. In our paper, the latter measure is used in line with Hu et al. (2017), Lee et al. (2018) and Tsai et al. (2020), and the results lend support to H1.3. Meanwhile, the impact of review total votes/likes is negative, which can be explained by the fact that old (less recent) reviews have more time to accumulate votes but are considered less helpful (H1.4). The negative effect of review total votes/likes on user-review affinity is consistent with the support for H1.4 in our data.

As regards the group of hypotheses on reviewer characteristics, reviewer expertise negatively affects user-review affinity, leading to H2.1 rejection. The negative sign remained unchanged even when we used Reviewer’s average number of (good) reviews and photos per day since his joining time (detailed results are provided on request).

Nevertheless, the interaction term between reviewer expertise and review variance has a positive effect on user-review affinity. This means that reviewer expertise moderates the relationship between review variance and user-review affinity, which is in line with the findings of Quaschning et al. (2015) and Filieri et al. (2019), where reviews written by expert reviewers are considered more helpful when deviating more from business average ratings (see Figure 3 where reviewer expertise (Expertise) and review variance (Variance) are standardized).

Figure 3. Interaction between review variance and reviewer expertise

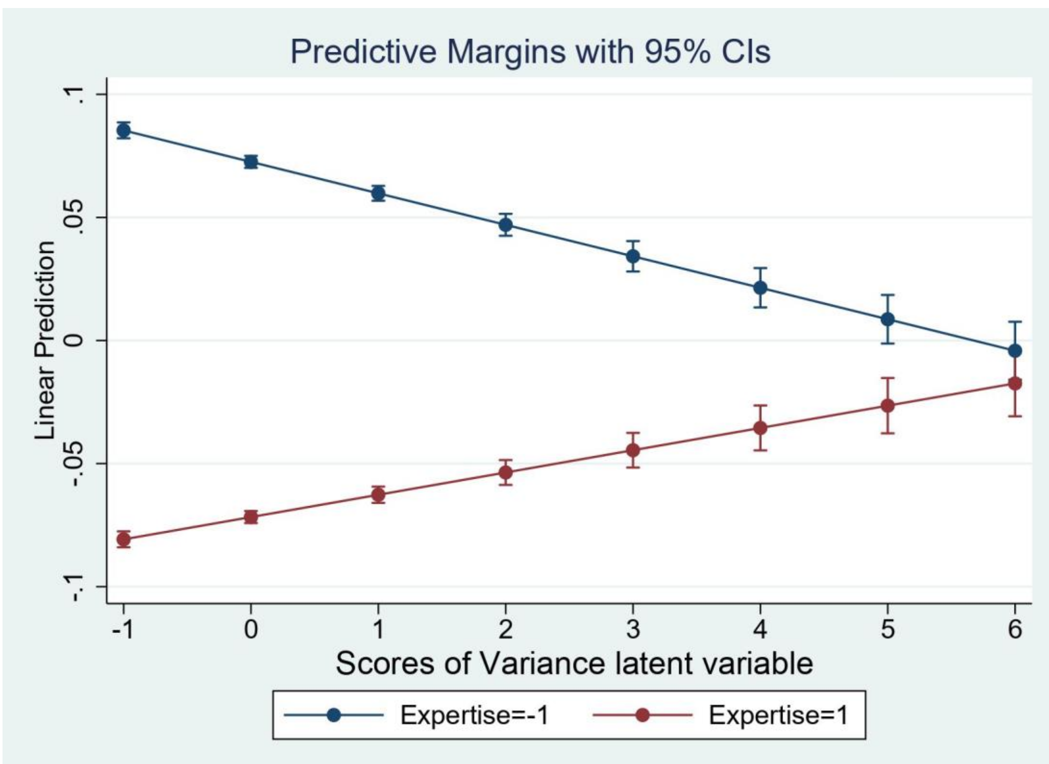
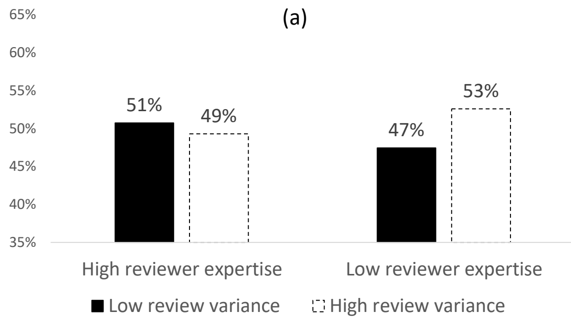
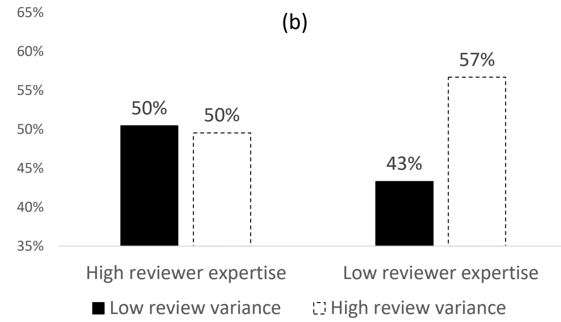
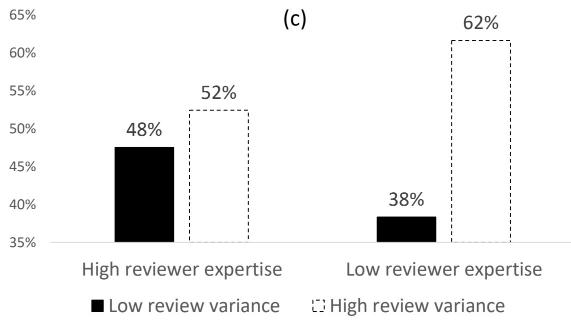


Figure 4. Review variance plotted by reviewer expertise





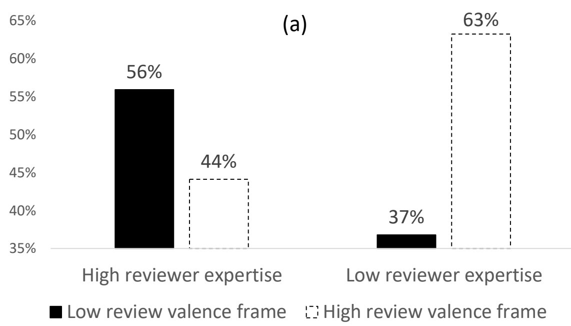


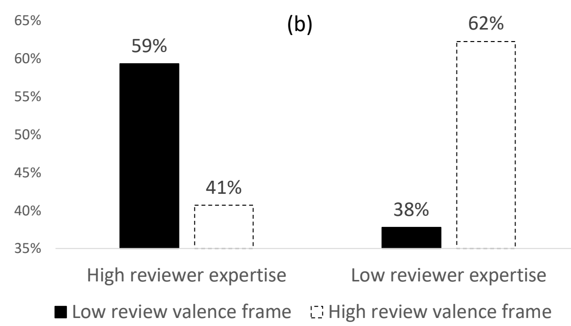
(a) Low = from the 10th percentile to the 25th percentile
High = from the 75th percentile to the 90th percentile

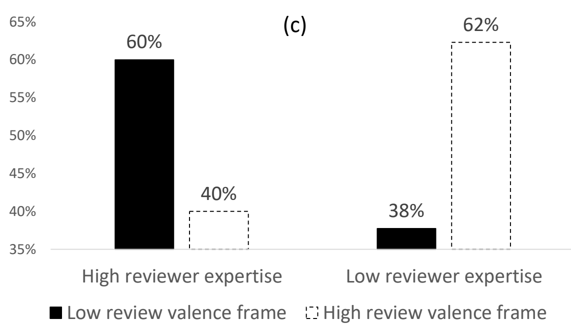
(b) Low = from the 5th percentile to the 25th percentile
High = from the 75th percentile to the 95th percentile

(c) Low = from the 1st percentile to the 25th percentile
High = from the 75th percentile to the 99th percentile

Figure 5. Review valence frame plotted by reviewer expertise







(a) Low = from the 10th percentile to the 25th percentile
High = from the 75th percentile to the 90th percentile

(b) Low = from the 5th percentile to the 25th percentile
High = from the 75th percentile to the 95th percentile

(c) Low = from the 1st percentile to the 25th percentile
High = from the 75th percentile to the 99th percentile

In our dataset, reviews written by reviewers with high expertise (reviewer expertise from the 75th percentile to the 90th, 95th, or 99th percentile) did not often deviate from business average ratings compared to their counterparts written by reviewers with low expertise (reviewer expertise from the 1^{st}, 5^{th}, or 10^{th} percentile to the 25^{th} percentile) as depicted in Figure 4. Nonetheless, when expert reviews deviated from business average ratings, they were often not in favor of the reviewed item (low review valence frame) vis-à-vis their nonexpert counterparts (see Figure 5) while high review valence frame increases user-review affinity (supported H1.1). This may explain why reviewer expertise correlated negatively with user-review affinity in our data.

The positive and statistically significant coefficient of reviewer social connectedness provides support for H2.2. This signifies that reviewers with many followers are likely to write high-quality and valuable reviews, which can prompt readers to cast helpfulness votes, leave comments, or re-read.

The positive effect of reviewer-user similarity on user-review affinity in Table 5 substantiates H2.3. This is in line with the findings of Neirotti et al. (2016) that when User and Reviewer know each other or share similar interests via prior interactions or common followships, they tend to trust the review and likely like, comment on, or re-read it.

In accordance with Yang et al. (2017), the positive coefficient of reviewer locality denotes that reviews written by local reviewers are deemed to trigger higher user-review affinity, corroborating H2.5. In addition, our data show that user-review affinity increases when User and Reviewer come from the same region, which was underexplored in the literature we consulted but can be explicated by high collectivism in Southeast Asia, where in-group members find shared values (Hofstede 2001).

As hypothesized, brand strength reduces user-review affinity and H3 is supported. Clearly, users might have already been familiar with the brand (Purnawirawan et al. 2015), so reviews may be considered less helpful. With regard to control variables, while review pictures were nonsignificant, review length correlates negatively with user-review affinity, signifying that long reviews are less likely to be voted for, commented on, or re-read.

With the tested conceptual model, we replicate the PLS-SEM algorithm presented by Venturini and Mehmetoglu (2019) in Python to compute the composite scores in Figure 2 for our predictive and prescriptive analytics. For robustness, we compare the results of the models with all variables, weighted (composite) scores and unweighted (composite) scales in the next subsection.

5. PREDICTIVE MODEL

In the previous section, we develop a set of independent variables and demonstrate that they significantly impact user affinity. In this section, we leverage these insights to develop a predictive model and demonstrate that our insights also apply to an out-of-sample instance. We first begin by considering machine learning algorithms that we will use for our predictive model.

According to Paul et al. (2018), Random Forest Classifier (RFC) is a widely-used ensemble learning algorithm to handle data imbalance. The upper bound to the generalization error of this algorithm is theoretically proven in the seminal paper of Breiman (2001) and its consistency is also confirmed in several recent papers with both theoretical analysis (e.g., Scornet et al. 2015, Wager and Athey 2018) and empirical studies (e.g., Calderoni et al. 2015, Mercadier and Lardy 2019). Many scholars find Random Forest's superior performance compared with other methods, such as regression tree and support vector machine (Wang et al. 2018d), logistic regression and artificial neural networks (ANN) (Wang et al. 2018b). Albeit outperformed by other techniques in some instances, Random Forest is still favored since it requires less parameter tuning (Ahmad et al. 2017, Mercadier and Lardy 2019). However, to select a robust model, we compare RFC with some other common algorithms in the literature and in practice (Abellán et al. 2017, Huber et al. 2019), namely, ANN (Hornik 1991), bagging classifier (BC) (Breiman 1996) and gradient boosting classifier (GBC) (Friedman 2001).

ANNs are also deemed effective for this classification problem (Aziz et al. 2018). Several techniques have been proposed to improve the performance of ANN (Lolli et al. 2017, Wang et al. 2018b, Huber et al. 2019). For instance, in Arcos-García et al.'s study (2017), the performance of their ANN model was not compromised by data imbalance while Huber et al.'s ANN algorithm (2019) can perform well in the presence of relaxed normality assumption provided that the dataset is big enough. Yet, ANNs are often considered a black box (Chen and Hao 2017) with many hyperparameters, e.g., number of layers and neurons, to fine-tune (Ahmad et al. 2017). In Wang et al.'s review (2018b), ANNs are suited to such specialized data domains as image and natural language processing, but outperformed by Random Forest in arbitrary domains.

In BC, the trees are built on randomly bootstrapped copies of the original instances, where features for node splitting can be drawn with or without replacement (Louppe and Geurts 2012). With this added randomness, the correlation between decision trees in the forest decreases and the model's performance is boosted, along with variance reduction and overfitting avoidance (Seyedhosseini and Tasdizen 2015, Mercadier and Lardy 2019). According to Scornet et al. (2015), BC is among the most computationally effective schemes for high-dimensional data.

GBC is a robust method to handle outliers and heterogenous attributes in multidimensional data (de Santis et al. 2017). The algorithm uses gradient-based approximations to split the tree node on the negative gradient for loss minimization (Athey et al. 2019), thereby enabling the optimization of an arbitrary loss function (Friedman 2001). Both BC and GBC are deemed effective for accuracy improvement of the classification problem (Dietterich 2000). Malik and Hussain's study (2018) is one of the earliest publications applying GBC for review helpfulness prediction based on review content variables and reviewer characteristics. Their comparative results show that GBC has lower (root) mean squared error than RFC and ANN.

In our paper, we bootstrap data within the six months before the beginning date of the first data point in the testing set to train the predictive models.

As regards some hyperparameters selected for our models, which are run on scikit-learn machine-learning package (Pedregosa et al. 2011), the three hyperparameters of interest in RFC are the number of decision trees (M), the tree depth and the number of features per tree ($mtry$). While the optimal number of features per tree receives a broad consensus in empirical findings (Wang et al. 2018b, Wang et al. 2018d), the number of decision trees and the tree depth vary across studies. For instance, the optimal number of nodes per tree is 5 in Tsagkrasoulis and Montana (2018), 8 in Zhou and Qiu (2018), 15 in Genuer et al. (2017) and Mercadier and Lardy (2019), and 20 in Chen et al. (2018). According to Ahmad et al. (2017), RFC performance deteriorates after the maximum depth exceeds 10, so we test the tree depth at 5, 8, and 10. We also run the scenarios where the tree depth is not limited ($treeDepth = \text{None}$). With respect to the number of trees (M), we try four thresholds, i.e., 30, 50, 100, and 200, to select the best one. To ensure fair comparisons, these hyperparameters are also applied to the models of BC, GBC, and ANN where appropriate. In particular, the ANN model has three hidden layer (M, 50, 15) for M equal to the number of decision trees.

After evaluating these predictive models based on precision, recall, and F1 scores, we select the ones with high performance to run prescriptive analytics, where reviews are rearranged as per the tested conceptual model. Particularly, reviews triggering higher estimated user-review affinity based on the conceptual model's parameters will be placed before those with lower user-review affinity. This new ranking (called *affinity-based ranking*) will replace the current review age-based ranking, and the trained predictive models will simulate if more users would like, comment on, or re-read the review within seven days.

We propose a predictive approach based on the selected features as described in the previous section for the platform to recommend reviews with high affinity to each user in a personalized fashion. In other words, our models are to predict if User will like, comment on, or re-read the review within seven days of reading. To evaluate the predictive models' performance on each month of the second half of year 2017, we bootstrapped the six months' data before the month under analysis to compile the training set and repeated this bootstrap-train-test procedure 30 times. Of a particular note is that there is no single model that outperformed others in all three criteria. The ANN method was the least stable with very large standard deviations compared to other models. BC, GBC, and RFC had similar performance, but the runtime of GBC was far longer. Hence, we will focus on discussing the BC and RFC results. The complete results of this evaluation exercise are available in the online appendix.

Overall, the models with weighted scores and unweighted scales yielded similar results, whereas the models with all variables are slightly better but their computation took more than double the runtime of their counterparts with dimensionality reduction. The only exception with respect to computational time is the RFC, where the processing time difference was only a few minutes. These results imply that the dimensionality reduction performed before was beneficial because the runtime savings were substantial while the predictive power loss was marginal.

Of particular note, we find that enlarging the training dataset by including less recent instances produced insignificant changes in the performance of RFC, BC, and GBC. Indeed, the models trained on the data one month before had comparable results to their counterparts trained on up to six months' data. This suggests that we can focus on a smaller yet more recent dataset to save the training time without compromising the predictive model's performance. This also lends empirical support for Zhang et al.'s (2017) statement that most recent data should be attended to.

Table 6. Confusion matrix for RFC averaged on monthly testing data (July–December)

Actual / Predicted	Positives			Negative		
Positives	356001.03 (0.19%)	351721.60 (0.21%)	353416.53 (0.19%)	30020.57 (0.98%)	30984.17 (1.03%)	31036.77 (0.95%)
Negatives	196430.97 (0.34%)	200710.40 (0.37%)	199015.47 (0.33%)	1755974.43 (0.02%)	1755010.83 (0.02%)	1754958.23 (0.02%)
	(1)	(2)	(3)	(1)	(2)	(3)

Note: in parentheses are the coefficients of variation. (1) model with all variables. (2) model with weighted scores. (3) model with unweighted scales. Number of estimators = 100. Max depth = None.

Table 7. Confusion matrix for BC averaged on monthly testing data (July–December)

Actual / Predicted	Positives			Negative		
Positives	357438.10 (0.31%)	350580.17 (0.27%)	351698.63 (0.27%)	36291.30 (1.19%)	37292.03 (1.12%)	36994.73 (1.13%)
Negatives	194993.90 (0.56%)	201851.83 (0.47%)	200733.37 (0.48%)	1749703.70 (0.02%)	1748702.97 (0.02%)	1749000.27 (0.02%)
	(1)	(2)	(3)	(1)	(2)	(3)

Note: in parentheses are the coefficients of variation. (1) model with all variables. (2) model with weighted scores. (3) model with unweighted scales. Number of estimators = 100.

Table 8. Predictive performance for RFC and BC averaged on monthly testing data (July–December)

	RFC			BC		
Precision	92.22% (0.07%)	91.90% (0.08%)	91.93% (0.07%)	90.78% (0.08%)	90.39% (0.10%)	90.48% (0.09%)
Recall	64.44% (0.97%)	63.67% (0.92%)	63.97% (0.95%)	64.70% (0.80%)	63.46% (0.80%)	63.66% (0.82%)
F1	75.87% (0.61%)	75.22% (0.59%)	75.44% (0.61%)	75.56% (0.51%)	74.57% (0.53%)	74.74% (0.54%)
	(1)	(2)	(3)	(1)	(2)	(3)

Note: in parentheses are the standard deviation. (1) model with all variables. (2) model with weighted scores. (3) model with unweighted scales. Number of estimators = 100. Max depth = None.

The confusion matrices in Table 6, Table 7, and Table 8 present the prediction results averaged over the latter half of year 2017 in our data. The BC models made more positive predictions, but their RBC counterparts made overall more true-positive (TP) predictions and less false-positive predictions, producing a higher precision. The F1 rates and forecast accuracy (TP + true negatives) of the RFC models were also higher. This might suggest that the recommender system based on RFC can work well for users who prefer to receive fewer yet more helpful reviews. Meanwhile, with BC, the system might boost the diversity of its recommendation. We will use RFC and BC for our prescriptive analysis for review recommendation.

6. PRESCRIPTIVE MODEL

In the previous section, we demonstrate that our predictive model can consistently predict user affinity for each review. Based on this insight, we next propose a prescriptive model where we recommend reviews based on the propensity for a review to be engaged by a user. We then use counterfactual simulations to show the performance of our personalized review commendation system.

First, to evaluate the effectiveness of different review recommendation strategies, we perform the counterfactual what-if analysis by using the inferences obtained from the predictive models (Dickerman and Hernán 2020), i.e., BC and RFC which yielded reliable predictive performance. As discussed in Section 4.2.1, our review platform arranges reviews in ascending order of review post time lapse, which is the original ranking in counterfactual what-if analysis. Reviews with ranking from 1st to 10th (first-page reviews) are considered promoted in our prescriptive analytics. However, since the platform can change this arrangement, we want to test if user-review affinity will increase if reviews are arranged in a personalized manner such that first-page reviews (reviews ranking 1st to 10th) are the most relevant or useful to the user concerned. More specifically, based on the confirmed conceptual model's parameters (Table 5) and a subset of data, we re-rank each review based on its estimated user-review affinity vis-à-vis other reviews (both read and unread) for the same business/item in descending order. Given that our model is particularly relevant for businesses with so many reviews that users may face information overload, our data subset for prescriptive analysis is focused on businesses with at least 50 reviews or an equivalent of five review pages as measured.

Note that, as can be seen in Table 9, there is no statistical difference in terms of business average rating, price range, and business age at the 1% level between the subsets of data where reviewed businesses had different thresholds for the minimum number of reviews by the end of 30 November 2017 (2017-11-30 23:59:59). All the statistics tests, namely t-test, KS test, and z-test, produced consistent results, and raising or lowering this threshold by 10 reviews did not change the statistical comparability of those subsets (see Table 9). The prescriptive analysis results reported in Table 10 and Table 11 are for businesses which had at least 50 reviews. In our prescriptive analytics, a business is deemed to have non-decreased user-review affinity when its positive user interactions (votes for, comments on, or re-reading of its reviews) simulated with affinity-based ranking are equal to or more than their counterparts with original ranking.

Table 9. p-value of statistics tests for subsets of the original and affinity-based ranking data

Businesses with / Indicators	≥ 40 reviews			≥ 50 reviews			≥ 60 reviews		
	t-test	KS test	z-test	t-test	KS test	z-test	t-test	KS test	z-test
Business average rating	0.320	0.789	0.321	0.480	0.974	0.480	0.615	0.999	0.615
Price range	0.093	0.774	0.092	0.209	0.208	0.118	0.513	0.999	0.513
Business age	0.535	0.908	0.534	0.789	0.985	0.789	0.879	0.973	0.879

Table 10. Proportion of simulated positives in subset of reviews ranked 1st–10th for businesses with at least 50 reviews

	Original ranking			Reranking		
RFC	12.70% (0.0009)	12.23% (0.0010)	12.07% (0.0012)	**31.60%** (0.0116)	**21.87%** (0.0165)	**21.09%** (0.0152)
BC	12.81% (0.0012)	12.99% (0.0014)	12.95% (0.0015)	**30.29%** (0.0158)	**29.42%** (0.0178)	**28.81%** (0.0190)
	(1)	(2)	(3)	(1)	(2)	(3)

Note: in parentheses are the standard deviations. (1) model with all variables. (2) model with weighted scores. (3) model with unweighted scales. In bold are the proportions which are statistically greater than their counterparts at the 1% significance level.

Table 11. Average positive user interaction rate for businesses with at least 50 reviews, reviews ranked 1st–10th under affinity-based re-ranking

	Original ranking			Reranking		
RFC	20.60% (0.2016)	20.00% (0.2000)	19.88% (0.2023)	**30.39%** (0.3229)	20.69% (0.2630)	20.06% (0.2567)
BC	20.30% (0.1931)	20.69% (0.2006)	20.66% (0.1994)	**29.15%** (0.3017)	**28.50%** (0.2930)	**27.71%** (0.2885)
	(1)	(2)	(3)	(1)	(2)	(3)

Note: in parentheses are the standard deviations. (1) model with all variables. (2) model with weighted scores. (3) model with unweighted scales. In bold are the proportions which are statistically greater than their counterparts at the 1% significance level.

As can be seen in Table 10 and Table 11, the reviews promoted by the novel affinity-based ranking, which is based on the conceptual model increased user affinity with the platform (by voting for, commenting on, or re-reading the reviews within seven days), and this improvement is statistically significant in most of the simulators considered at the 1% significance level. In particular, businesses with at least 50 reviews, the re-ranking increased user interactions in all simulators. Users who read the promoted reviews of those businesses also interacted more with the reviews and that increase in interactions was statistically significant at the 1% level in most simulators. Thus, review platforms can leverage this insight to arrange product reviews in a personalized fashion for each user to boost user-review affinity. Moreover, in line with Table 6 and Table 7, the BC-based system produced more recommendations and thus likely boost the diversity of recommended reviews.

7. DISCUSSIONS AND CONCLUSIONS

Online reviews have increasingly become an integral part of many online platforms in recent years. While these platforms spend considerable resources and efforts to attract users to contribute online reviews, they also face a critical issue where their consumers have too many reviews to read, leading them to suffer from information fatigue. In this paper, we propose a framework to alleviate such an issue. Specifically, we develop a personalized review recommendation system that can help platforms selectively displays reviews to their users based on the propensity that each user is going to engaged with each review.

Our framework begins by conducting an exploratory analysis where we extensively survey prior works to identify key independent variables that could affect user affinity (i.e., the tendency that a user would like, comment, or re-read the reviews). We also conduct factory analysis and confirmatory composite analysis to reduce the complexity of the variables and avoid multi-collinearity. Based on this exercise, we confirm several important features studied in the literature, hypothesize the potential impact of these variables, and statistically test our hypotheses.

Following that, we leverage the insights uncovered from our exploratory model to develop a predictive model. Here, our objective is to ensure that our insights apply to out-of-sample instances to verify the external validity of our finding. In addition, this exercise also allows us to predict the propensity that each user would interact with each review, which is the key ingredient that we use to develop the personalized review recommendation system in the next step. Overall, we find that off-the-shelf machine learning classification algorithms can consistently predict the level of user affinity based on the predictors that we derive from our exploratory analysis.

With a consistently accurate predictive model, we proceed to develop a prescriptive framework where we re-rank reviews based on their potential user affinity. We evaluate this personalized review recommendation system using counterfactual simulations where we demonstrate that re-ranking reviews can attain significantly more engagements from users, which generally lead to higher user satisfaction and retention with the platforms.

REFERENCES

Abellán J, Mantas CJ, Castellano JG (2017) A Random Forest approach using imprecise probabilities. *Knowledge-Based Systems* 134:72–84.

Aghakhani N, Oh O, Gregg DG, Karimi J (2021) Online Review Consistency Matters: An Elaboration Likelihood Model Perspective. *Information Systems Frontiers* 23(5):1287–1301.

Ahmad MW, Mourshed M, Rezgui Y (2017) Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings* 147:77–89.

Akter S, Fosso Wamba S, Dewan S (2017) Why PLS-SEM is suitable for complex modelling? An empirical illustration in big data analytics quality. *Production Planning & Control* 28(11-12):1011–1021.

Arcos-García Á, Soilán M, Álvarez-García JA, Riveiro B (2017) Exploiting synergies of mobile mapping sensors and deep learning for traffic sign recognition systems. *Expert Systems with Applications* 89:286–295.

Armentano MG, Godoy D, Amandi AA (2013) Followee recommendation based on text analysis of micro-blogging activity. *Information Systems* 38(8):1116–1127.

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *The Annals of Statistics* 47(2):1148–1178.

Aziz R, Verma CK, Srivastava N (2018) Artificial Neural Network Classification of High Dimensional Data with Novel Optimization Approach of Dimension Reduction. *Annals of Data Science* 5(4):615–635.

Bai X, Barla Cambazoglu B, Gullo F, Mantrach A, Silvestri F (2017) Exploiting search history of users for news personalization. *Information Sciences* 385-386:125–137.

Blal I, Sturman MC (2014) The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hospitality Quarterly* 55(4):365–375.

Bobko P, Roth PL, Buster MA (2007) The Usefulness of Unit Weights in Creating Composite Scores: A Literature Review, Application to Content Validity, and Meta-Analysis. *Organizational Research Methods* 10(4):689–709.

Breiman L (1996) Bagging predictors. *Machine learning* 24(2):123–140.

--- (2001) Random Forests. *Machine Learning* 45(1):5–32.

Calderoni L, Ferrara M, Franco A, Maio D (2015) Indoor localization in a hospital environment using Random Forest classifiers. *Expert Systems with Applications* 42(1):125–134.

Chen M, Wang X, Feng B, Liu W (2018) Structured random forest for label distribution learning. *Neurocomputing* 320:171–182.

Chen Y-J, Dai T, Korpeoglu CG, Körpeoğlu E, Sahin O, Tang CS, Xiao S (2020) OM Forum—Innovative Online Platforms: Research Opportunities. *Manufacturing & Service Operations Management* 22(3):430–445.

Chen Y, Hao Y (2017) A feature weighted support vector machine and K-nearest neighbor algorithm for

stock market indices prediction. *Expert Systems with Applications* 80:340–355.
Chevalier JA, Mayzlin D (2006) The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* 43(3):345–354.
Chin WW, Marcolin BL, Newsted PR (2003) A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research* 14(2):189–217.
Chua AYK, Banerjee S (2017) Analyzing review efficacy on Amazon.com: Does the rich grow richer? *Computers in Human Behavior* 75:501–509.
de Santis RB, de Aguiar EP, Goliatt L (2017) Predicting material backorders in inventory management using machine learning. *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (IEEE), 1–6.
Dickerman BA, Hernán MA (2020) Counterfactual prediction is not only for causal inference. *European Journal of Epidemiology* 35(7):615–617.
Dietterich TG (2000) An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning* 40(2):139–157.
Dunn TJ, Baguley T, Brunsden V (2014) From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology* 105(3):399–412.
Eirinaki M, Gao J, Varlamis I, Tserpes K (2018) Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions. *Future Generation Computer Systems* 78:413–418.
Fang B, Ye Q, Kucukusta D, Law R (2016) Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management* 52:498–506.
Filieri R, Alguezaui S, McLeay F (2015) Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management* 51:174–185.
Filieri R, Raguseo E, Vitari C (2018) When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior* 88:134–142.
--- (2019) What moderates the influence of extremely negative ratings? The role of review and reviewer characteristics. *International Journal of Hospitality Management* 77:333–341.
Filieri R, Vitari C, Raguseo E (2021) Extremely Negative Ratings and Online Consumer Review Helpfulness: The Moderating Role of Product Quality Signals. *Journal of Travel Research* 60(4):699–717.
Fleder D, Hosanagar K (2009) Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* 55(5):697–712.
Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18(1):39–50.
Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29(5):1189–1232.
Gao B, Hu N, Bose I (2017) Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems* 95:1–11.
Genuer R, Poggi J-M, Tuleau-Malot C, Villa-Vialaneix N (2017) Random Forests for Big Data. *Big Data Research* 9:28–46.
Gonzalez Camacho LA, Alves-Souza SN (2018) Social network data to alleviate cold-start in recommender system: A systematic review. *Information Processing & Management* 54(4):529–544.
Gorgoglione M, Panniello U, Tuzhilin A (2019) Recommendation strategies in personalization applications. *Information & Management* 56(6, Article 103143).
Hair JF, Howard MC, Nitzl C (2020) Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research* 109:101–110.
Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019) When to use and how to report the results of PLS-SEM. *European Business Review* 31(1):2–24.
Hall RJ, Snell AF, Foust MS (1999) Item Parceling Strategies in SEM: Investigating the Subtle Effects of

Unmodeled Secondary Constructs. *Organizational Research Methods* 2(3):233–256.
Henseler J, Hubona G, Ray Pauline A (2016) Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems* 116(1):2–20.
Ho-Dac NN, Carson SJ, Moore WL (2013) The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing* 77(6):37–53.
Hofstede GH (2001) *Cultures consequences: comparing values, behaviors, institutions, and organizations across nations*, 2nd ed. (Sage, Thousand Oaks, Calif).
Hong H, Xu D, Wang GA, Fan W (2017) Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decision Support Systems* 102:1–11.
Hornik K (1991) Approximation capabilities of multilayer feedforward networks. *Neural Networks* 4(2):251–257.
Hosanagar K, Fleder D, Lee D, Buja A (2014) Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science* 60(4):805–823.
Hu N, Liu L, Zhang JJ (2008) Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management* 9(3):201–214.
Hu X, Yang Y (2021) What makes online reviews helpful in tourism and hospitality? a bare-bones meta-analysis. *Journal of Hospitality Marketing & Management* 30(2):139–158.
Hu Y-H, Chen K (2016) Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management* 36(6):929–944.
Hu Y-H, Chen K, Lee P-J (2017) The effect of user-controllable filters on the prediction of online hotel reviews. *Information & Management* 54(6):728–744.
Huang C, Jiang W, Wu J, Wang G (2020) Personalized Review Recommendation based on Users' Aspect Sentiment. *ACM Transactions on Internet Technology* 20(4, Article 42).
Huang N, Hong Y, Burtch G (2017) Social Network Integration and User Content Generation: Evidence from Natural Experiments. *MIS quarterly* 41(4):1035–1058.
Huber J, Müller S, Fleischmann M, Stuckenschmidt H (2019) A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research* 278(3):904–915.
Ji P, Wayne Fu W (2013) Love Internet, love online content: Predicting Internet affinity with information gratification and social gratifications. *Internet Research* 23(4):396–413.
Karimi S, Wang F (2017) Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems* 96:39–48.
Khern-am-nuai W, Kannan K, Ghasemkhani H (2018) Extrinsic versus Intrinsic Rewards for Contributing Reviews in an Online Platform. *Information Systems Research* 29(4):871–892.
Lee M, Jeong M, Lee J (2017) Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management* 29(2):762–783.
Lee P-J, Hu Y-H, Lu K-T (2018) Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics* 35(2):436–445.
Lee S, Choeh JY (2016) The determinants of helpfulness of online reviews. *Behaviour & Information Technology* 35(10):853–863.
Lee Y-J, Hosanagar K, Tan Y (2015) Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science* 61(9):2241–2258.
Li L (2018) Sentiment-enhanced learning model for online language learning system. *Electronic Commerce Research* 18(1):23–64 journal article.
Li S, Cheng X, Su S, Jiang L (2016) Followee Recommendation in Event-Based Social Networks. Gao H, Kim J, Sakurai Y, eds. *Database Systems for Advanced Applications* (Springer International Publishing), 27–42.
Li Z, Fang X, Bai X, Sheng ORL (2017) Utility-Based Link Recommendation for Online Social Networks. *Management Science* 63(6):1938–1952.
Liang S, Schuckert M, Law R (2019) How to improve the stated helpfulness of hotel reviews? A multilevel

approach. *International Journal of Contemporary Hospitality Management* 31(2):953–977.
Lin M, Lucas Jr HC, Shmueli G (2013) Research commentary—too big to fail: large samples and the p-value problem. *Information Systems Research* 24(4):906-917.
Liu Y, Han L, Gou Z, Yang Y (2019) Personalized Recommendation via Trust-Based Diffusion. *IEEE Access* 7:94195–94204.
Lolli F, Gamberini R, Regattieri A, Balugani E, Gatos T, Gucci S (2017) Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics* 183:116–128.
Louppe G, Geurts P (2012) Ensembles on random patches. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Springer), 346–361.
Lu S, Wu J, Tseng S-L (2018) How Online Reviews Become Helpful: A Dynamic Perspective. *Journal of Interactive Marketing* 44:17–28.
Luca M, Zervas G (2016) Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Management Science* 62(12):3412–3427.
Ma Y, Xiang Z, Du Q, Fan W (2018) Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management* 71:120–131.
Malgonde O, Zhang H, Padmanabhan B, Limayem M (2020) Taming the Complexity in Search Matching: Two-Sided Recommender Systems on Digital Platforms. *MIS Quarterly* 44(1):48–84.
Malik MSI, Hussain A (2018) An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing & Management* 54(1):88–104.
Marquaridt DW (1970) Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics* 12(3):591–612.
Mason CH, Perreault WD (1991) Collinearity, Power, and Interpretation of Multiple Regression Analysis. *Journal of Marketing Research* 28(3):268–280.
Mathias M, Timofte R, Benenson R, Van Gool L (2013) Traffic sign recognition — How far are we from the solution? *2013 International Joint Conference on Neural Networks (IJCNN)*, 1–8.
Mercadier M, Lardy J-P (2019) Credit spread approximation and improvement using random forest regression. *European Journal of Operational Research* 277(1):351–365.
Mo J, Sarkar S, Menon S (2018) Know When to Run: Recommendations in Crowdsourcing Contests. *MIS Quarterly* 42(3):919–944.
Mudambi SM, Schuff D (2010) Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly* 34(1):185–200.
Neirotti P, Raguseo E, Paolucci E (2016) Are customers' reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning. *International Journal of Information Management* 36(6):1133–1143.
O'Mahony MP, Smyth B (2010) A classification-based review recommender. *Knowledge-Based Systems* 23(4):323–329.
Orlikowski WJ, Scott SV (2014) What Happens When Evaluation Goes Online? Exploring Apparatuses of Valuation in the Travel Sector. *Organization Science* 25(3):868–891.
Paul A, Mukherjee DP, Das P, Gangopadhyay A, Chintha AR, Kundu S (2018) Improved Random Forest for Classification. *IEEE Transactions Image Processing* 27(8):4012–4024.
Paul D, Sarkar S, Chelliah M, Kalyan C, Sinai Nadkarni PP (2017) Recommendation of High Quality Representative Reviews in e-commerce. Proceedings of the Eleventh ACM Conference on Recommender Systems, Association for Computing Machinery.
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12(85):2825–2830.
Perse EM (1986) Soap opera viewing patterns of college students and cultivation. *Journal of Broadcasting & Electronic Media* 30(2):175–193.
Prawesh S, Padmanabhan B (2014) The "Most Popular News" Recommender: Count Amplification and Manipulation Resistance. *Information Systems Research* 25(3):569–589.
Purnawirawan N, Eisend M, De Pelsmacker P, Dens N (2015) A Meta-analytic Investigation of the Role of

Valence in Online Reviews. *Journal of Interactive Marketing* 31:17–27.
Qian X, Feng H, Zhao G, Mei T (2014) Personalized Recommendation Combining User Interest and Social Circle. *IEEE Transactions on Knowledge and Data Engineering* 26(7):1763–1777.
Qiao D, Lee S-Y, Whinston AB, Wei Q (2020) Financial Incentives Dampen Altruism in Online Prosocial Contributions: A Study of Online Reviews. *Information Systems Research* 31(4):1361–1375.
Quaschning S, Pandelaere M, Vermeir I (2015) When Consistency Matters: The Effect of Valence Consistency on Review Helpfulness. *Journal of Computer-Mediated Communication* 20(2):136–152.
Rao R, Plotnick L, Hiltz SR (2017) Supporting the Use of Social Media by Emergency Managers: Software Tools to Overcome Information Overload. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 304–312.
Salehan M, Zhang S, Aghakhani N (2017) A recommender system for restaurant reviews based on consumer segment. Americas Conference on Information Systems (AMCIS).
Scornet E, Biau G, Vert J-P (2015) Consistency of random forests. *The Annals of Statistics* 43(4):1716–1741.
Sethi V, King WR (1994) Development of measures to assess the extent to which an information technology application provides competitive advantage. *Management Science* 40(12):1601–1627.
Seyedhosseini M, Tasdizen T (2015) Disjunctive normal random forests. *Pattern Recognition* 48(3):976–983.
Siering M, Muntermann J, Rajagopalan B (2018) Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems* 108:1–12.
Sijtsma K (2009) On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika* 74(1):107–120.
Sinai A (2004) Preemption, Changing Structure, and U. S. Monetary Policy. *American Economic Review* 94(2):49–52.
Sivasubramaniam J, Chandrasekar C (2019) Proximal spatial vector and affinity coefficient for multimodal biometric secured social network communication. *World Wide Web* 22(4):1751–1764.
Son LH (2015) HU-FCF++: A novel hybrid method for the new user cold-start problem in recommender systems. *Engineering Applications of Artificial Intelligence* 41:207–222.
Sparks BA, Browning V (2011) The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management* 32(6):1310–1323.
Sridhar S, Srinivasan R (2012) Social Influence Effects in Online Product Ratings. *Journal of Marketing* 76(5):70–88.
Sun M (2012) How Does the Variance of Product Ratings Matter? *Management Science* 58(4):696–707.
Tsagkrasoulis D, Montana G (2018) Random forest regression for manifold-valued responses. *Pattern Recognition Letters* 101:6–13.
Tsai C-F, Chen K, Hu Y-H, Chen W-K (2020) Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management* 80(Article 104122).
Tsao H-Y, Chen M-Y, Lin H-CK, Ma Y-C (2019) The asymmetric effect of review valence on numerical rating: A viewpoint from a sentiment analysis of users of TripAdvisor. *Online Information Review* 43(2):283–300.
Tsao W-C, Hsieh M-T, Shih L-W, Lin TMY (2015) Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management* 46:99–111.
Venturini S, Mehmetoglu M (2019) plssem: A Stata Package for Structural Equation Modeling with Partial Least Squares. *Journal of Statistical Software* 88(8):1–35 factor analysis; latent variables; partial least squares; PLS; PLS-PM; PLS-SEM; path models; Stata; structural equation modeling; SEM.
Wager S, Athey S (2018) Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* 113(523):1228–1242.
Wang C, Zhang X, Hann I-H (2018a) Socially Nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings. *Information Systems Research* 29(3):641–655.
Wang S, Aggarwal C, Liu H (2018b) Random-Forest-Inspired Neural Networks. *ACM Transactions on*

Intelligent Systems and Technology 9(6, Article 69).
Wang Y, Xia S-T, Tang Q, Wu J, Zhu X (2018c) A Novel Consistent Random Forest Framework: Bernoulli Random Forests. *IEEE Transactions on Neural Networks and Learning Systems* 29(8):3510–3523.
Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S (2018d) Random Forest based hourly building energy prediction. *Energy and Buildings* 171:11–25.
Wixted JT, Ebbesen EB (1997) Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition* 25(5):731–739.
Wu J (2017) Review popularity and review helpfulness: A model for user review effectiveness. *Decision Support Systems* 97:92–103.
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. *Tourism Management* 58:51–65.
Xu L, Duan JA, Hu YJ, Cheng Y, Zhu Y (2019) Forward-Looking Behavior in Mobile Data Consumption and Targeted Promotion Design: A Dynamic Structural Model. *Information Systems Research* 30(2):616–635.
Xu Y, Zhou M, Han S (2015) Feature representation for microblog followee recommendation in classification framework. *2015 Seventh International Conference on Advanced Computational Intelligence (ICACI)*, 318–322.
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.
Yang S, Zhou Y, Yao J, Chen Y, Wei J (2019) Understanding online review helpfulness in omnichannel retailing. *Industrial Management & Data Systems* 119(8):1565–1580.
Yong AG, Pearce S (2013) A Beginners Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology* 9(2):79–94.
Yu Y, Gao Y, Wang H, Wang R (2018) Joint user knowledge and matrix factorization for recommender systems. *World Wide Web* 21(4):1141–1163.
Yu Y, Khern-am-nuai W, Pinsonneault A, Wei Z (2020) Social Interactions and Bias in the Evaluation of Online Reviews. *Available at SSRN 3369407*.
Zhang F, Liu Q, Zeng A (2017) Timeliness in recommender systems. *Expert Systems with Applications* 85:270–278.
Zhang L, Ma B, Cartwright DK (2013) The impact of online user reviews on cameras sales. *European Journal of Marketing* 47(7):1115–1128.
Zhang S, Salehan M, Leung A, Cabral I, Aghakhani N (2018) A recommender system for cultural restaurants based on review factors and review sentiment. *Americas Conference on Information Systems (AMCIS)*.
Zhang Y, Lin Z (2018) Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications* 27:1–10.
Zhou S, Guo B (2017) The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems* 93:77–87.
Zhou Y, Qiu G (2018) Random forest for label ranking. *Expert Systems with Applications* 112:99–109.