A FRAMEWORK FOR AFFINITY-BASED PERSONALIZED REVIEW RECOMMENDATION

Abstract: Online review platforms have proliferated thanks to technological advances and consumers' increased dependence on each other's opinions in purchase decisions. However, users typically face an enormous number of online reviews and suffer from information overload. Unlike existing studies that rely mainly on popularity, crowd-based evaluation, or filtering methods, we propose a framework 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 he/she will like (hit the helpfulness vote/like button), comment on, or re-read those reviews, whereby user login time increases, which in turn correlates positively with user affinity toward the platform. We hypothesize a conceptual model, conduct predictive analytics, and perform counterfactual simulations on the log data of a large restaurant review platform in Southeast Asia and find that reviewer-user similarity is among the most significant explanatory factors, which is in line with the Asian culture. Built on the results of the explanatory analysis, machine learning-based predictive models are then applied to predict the likelihood that each user will interact with each review for each business. Our counterfactual analysis demonstrates the potential of the resultant affinity-based ranking to increase user engagement with the platform.

Keywords: review recommendation, user affinity, online platform, service operations

1. INTRODUCTION

Online review platforms have become one of the primary data sources for consumers (Siering and Janze 2019). Prior studies have shown that many consumers rely on online reviews in their decision-making process (Choi et al. 2022), resulting in strong empirical connections between online reviews and product sales (Sun 2012, Anderson and Lawrence 2014). Hence, many platforms that aggregate online reviews have proliferated (Luca and Zervas 2016), especially for products that consumers cannot directly evaluate before consumption such as restaurants and hotels (Khern-am-nuai et al. 2018). These platforms have competed to attract and retain content contributors over the years (Qiao et al. 2021). However, content increases also trigger unforeseen issues 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 likely causes difficulty in filtering pertinent information (Zhou and Guo 2017). This problem has become increasingly important since recent studies have reported that it may dampen platforms' success (Chen et al. 2020). Several approaches have been adopted by online review platforms to mitigate this issue, including the utilization of crowd-based content assessment such as review (un)helpfulness scores (Orlikowski and Scott 2014) and structured content filtering such as tags and badges (Rao et al. 2017). This research studies an approach to help users overcome information overload with online reviews. Specifically, we build a personalized review recommendation framework on a theory-driven yet readily implementable model.

In the era of big data, the concept of personalized recommendation has been widely adopted by various services. Service providers, including online platforms, have invested in capturing and analyzing data on customers' digital trails of activities, such as browsing history, geographical locations, purchases, likes, and comments, to customize their service offerings/delivery such that customer satisfaction and profitability are improved (Cohen 2018, Caro et al. 2020). Indeed, online platforms have transformed a significant part of service operations given that they can collect data on customers' tastes, habits, and social networks to make appropriate recommendations (Cohen 2018). The literature has covered the design and use of personalized recommendation in multiple contexts, e.g., 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 fairly rare in research and practice. Meanwhile, existing works that 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), review quality (Paul et al. 2017), and consumer segment (Salehan et al. 2017). Inspired by this gap in research and practice, this paper leverages a unique dataset obtained via collaboration with a large restaurant review platform in Southeast Asia to propose a personalized review recommendation framework built on user-review affinity, which can be broadly characterized as users' positive attachment to media content (Ji and Fu 2013). In effect, emotional attachment can boost customer loyalty (Khan and Rahman 2017). We discuss the proxies for user-review affinity adopted herein in Section 3.2.

Figure 1. Research design

Our research design includes exploratory, predictive, and counterfactual analyses (see Figure 1). First, we survey prior online review studies to identify factors that potentially impact user-review affinity. Particularly, this article runs exploratory analysis by partial least squares structural equation modeling or PLS-SEM (Henseler et al. 2016), which is commonly performed for hierarchical models in big data analytics (Akter et al. 2017), to test a conceptual model that connects explanatory variables to user-review affinity. Afterward, we develop a predictive model using machine learning (ML) algorithms and factors identified in the previous step as predictors of the likelihood that a user will interact with each review for each business. To test the efficacy of the proposed review recommendation framework in our exploratory analysis, we rely on the results from our predictive modeling to run counterfactual simulations, which have gained traction as a research methodology to validate proposed strategies (Xu et al. 2019).

This paper's contributions are threefold. First, we identify and verify crucial 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 ML models built on the verified factors can achieve comparable predictive performance at shorter runtimes vis-à-vis their high-dimensional counterparts. Lastly, we illustrate that arranging reviews in descending order of their predicted effect on user affinity rather than in time order is more effective in improving user affinity.

The rest of the paper is organized as follows. In Section 2, we review background literature related to this study. Our data and research context are described in Section 3. Section 4 discusses our exploratory model, followed by predictive modeling in Section 5 and counterfactual analysis in Section 6. Finally, we discuss limitations and future research avenues and conclude our study in Section 7.

2. RELATED LITERATURE

In this section, we survey background literature that is related to our study. Particularly, we first review research on recommender systems and then discuss previous works on 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 purchase history and preferences (Xu et al. 2017, 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 platform users' demographics, other details to build the model can derive from their own comments, search history (Bai et al. 2017), or social networks (Li et al. 2017). Gonzalez Camacho and Alves-Souza (2018) find that social networks parlayed in collaborative algorithms are useful to give recommendations to new users or those with incomplete profiles, where preferences are not specified, or to suggest new items to existing users, who may 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 his/her 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 items in peers' profiles may not perfectly fit the user in question, which requires diversifying the algorithm into identifying or exploring items that are probably better suited for 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 as 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 (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 profit, are considered (Gorgoglione et al. 2019). As recommender systems aim to help customers improve experiences and interactions with businesses, which consist of browsing, purchasing, and giving feedback (Gorgoglione et al. 2019), our article focuses on predicting the probability that a user in question will interact more with a given review for a certain business via liking (hitting the helpfulness vote/like button), commenting on, or re-reading (by clicking again on) that review. This will be further justified and elaborated 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 customers' 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 efficacy for sales conversion, review popularity is as vital as review helpfulness, which emphasizes the relevance to the customer under analysis. Also, as the country, where the online review platform in our research is based, scores high on collectivism (Hofstede 2001), implying a strong inclination for conformity (Tsao et al. 2015), review popularity in collaborative-based recommender systems can be relevant.

With user tastes and preferences evolving over time (Zhang et al. 2017), recently posted reviews are regarded by review readers as more helpful (Hu et al. 2008, Filieri et al. 2015, Zhou and Guo 2017). Several other features of the review itself and its reviewer are also found significant in the helpfulness of the review and its impact on sales (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 pertinent publications. Further, Hong et al. (2017) validate the moderating role of the platform host and product category in determining the helpfulness of the review. In effect, consumers deem reviews obtained from third-party platforms to be more reliable than those from seller-hosted platforms, and experience products/services, whose quality evaluation is subjective and user-specific and thus hard to obtain via objective information search, necessitate consulting more online reviews (Anderson and Lawrence 2014, Mankad et al. 2016), especially those whose reviewers have common interests and personalities with the customer in question (Hong et al. 2017).

Inasmuch as our dataset was collected from a third-party platform for online reviews on hospitality businesses, which primarily provide experience products, we can focus our model on attributes associated with reviews, review writers (reviewers), and (platform) users (who are seeking reviews).

2.2 Related Theories

In addition to Hofstede's (2001) cultural dimensions, which have been widely utilized across disciplines, including service operations (Yayla-Küllü et al. 2015), our work also adopts the network effect (Farrell and Klemperer 2007, Chen et al. 2020) and signaling theory (Spence 1973). In effect, the platform with most users is valued most (Chen et al. 2020) as customers of a product/service/system value compatibility with peers (Farrell and Klemperer 2007). Likewise, reviews with most likes are most probably considered helpful and reviewers who have received many helpfulness votes from platform users or have a large network of followers and followees are likely to write helpful reviews. This may well be supported in collectivist culture as in Southeast Asia, where conformity to the norm is appreciated (Hofstede 2001). As regards signaling theory, consumers, given asymmetric information, have to use observable cues, aka signals, to evaluate product or service quality (Spence 2002, Filieri et al. 2021). In our context, attributes of a review and its reviewer can be employed as signals to platform users. These underpinnings will be elaborated in the next section, where potential variables for our model are operationalized and hypothesized. By building our framework on theoretical foundations, we respond to the need for more data-driven service operations models that are also theoretically grounded (Huang and Rust 2013), and ensure that the findings are not confined to our specific dataset but likely generalizable (see Bansal et al. 2020).

It should be noted that a majority of existing publications focus on English reviews or English-based contexts as highlighted by Zhang and Lin (2018). The literature reviews of Gao et al. (2017) and Wu (2017) illustrate that the review platforms frequently studied are headquartered in the U.S., e.g., Yelp, Amazon, TripAdvisor, Apple's App Store, Yahoo!, and CNET. Zhang and Lin (2018) therefore argue that models developed in English-based contexts may not be perfectly transferable to non-English settings. By leveraging the features substantiated in the literature, we can test if research results for review recommendation in developed, western, or English-speaking nations are applicable for Asian markets, where Hofstede's (2001) cultural dimensions can differ.

3. DATA AND RESEARCH CONTEXT

3.1 Data Descriptions

We obtained our data through collaboration with a large restaurant review platform in Southeast Asia, which then had over three million users and more than ten million reviews and photos for restaurants and other businesses (e.g., beauty salons and shopping malls) in around three hundred thousand locations in its home market.

Table 1. Descriptive statistics of the dataset

Our data contain 4,151,904 user-review interactions on the platform in 2017, including those where users read, liked, or commented on reviews. For 216,556 unique (platform) users, 435,512 reviews, 57,218 reviewers (review writers), and 76,703 businesses in the data, descriptive statistics of interactions on the platform are given in Table 1. Users within the 5^{th}–95^{th} percentile had, on average, 8.47 interactions with reviews on the platform. Nonetheless, most of them had no more than four interactions, suggesting that the data are heavily left-skewed. Meanwhile, on the review side, each review received 5.45 interactions from platform users on average, but the majority of them attained no more than three interactions. Reviewers on the platform had their reviews read/liked/commented on 9.36 times on average. Businesses on the platform, on the other hand, had their reviews read/liked/commented on 23.78 times on average. The left skewness nature of the interactions is observed at these levels too.

3.2 User-Review Affinity

The dependent variable of this study is user-review affinity, a variable of interest to most online review platforms as it correlates strongly with media-viewing time and frequency (Perse 1986, Ji and Fu 2013), which are directly associated with the platform's revenue and sustainability. To operationalize the concept of user-review affinity we consider three activities: like (hit the helpfulness vote/like button), comment, and re-read (by clicking again on the review) as indicators of increased user-review affinity (hereafter referred to as *user-review affinity*). Specifically, user affinity for a review equals 1 if, within seven days after the initial read, the user liked (hit the helpfulness vote or like button), re-read (by clicking again on), or commented on the review, and 0 otherwise. We select the seven-day threshold as the Ebbinghaus forgetting curve is relatively flat afterward (Wixted and Ebbesen 1997, Li 2018).

From platform users' perspective, these activities indicate their affinity for reviews. In effect, helpfulness votes indicate that users find the review helpful for their decision-making (Tsai et al. 2020, Filieri et al. 2021), which in turn increases their affinity toward the review and the platform. In addition, prior literature has shown that helpful reviews tend to receive reader comments (Malik and Hussain 2018). As such, prompting review readers to make comments can theoretically herald as review relevance and increased user affinity since they log in more often or longer. Lastly, prior works also consider readership for review evaluation (Chua and Banerjee 2017). As users re-read (by clicking again on) the review within seven days, they would visit the platform more often and spend more time on the content, which can also be considered increased affinity. Additionally, from the platform's perspective, these activities are also regarded by our collaborating platform as a key performance measure because users tend to stay longer and access the platform more often when they like, comment on, or re-read reviews. These activities are also in line with the construct of user affinity in the literature (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 from users (i.e., reviews that users are likely to like, comment on, or re-read). We next develop a framework to produce personalized review recommendations based on user-review affinity. In the next section, we begin by exploring prior literature on factors that may affect user-review affinity.

4. EXPLORATORY MODEL

In this section, we describe our exploratory model that is built to identify factors that could influence user-review affinity. In this regard, we draw on prior works that study the effect of numerous variables on user affinity. Results from this exploratory model will be used to inform our predictive modeling and counterfactual analysis.

Here, we frame the problem at hand as a classification problem (i.e., the target variable captures whether a user would like/comment on/re-read a review or not). We follow Mathias et al.'s (2013) three fine-grained steps for a classification model, i.e., feature extraction, dimensionality reduction, and classification. User (she) and Reviewer (he) denote the focal (platform) user (review reader) and reviewer (review writer), 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 model belong to three groups, namely reviewer characteristics, review features, and product attributes. Based on an extensive survey of previous research, we compile in Table 2 the list of explanatory variables used in this study.

Table 2. List of variables

We next develop an integrated model that verify the influence of Table 2's variables on user affinity. Since there are 46 variables, we face two issues. First, as demonstrated by Lin et al. (2013), having many exploratory variables with millions of observations would most likely result in an overfitted model. Second, including too many input variables can cause computational issues. 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 scholarship shows 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 that relevant variables are incorporated while multicollinearity is avoided, we first conduct exploratory factor analysis (EFA). Eigenvalues and factor rotation (Yong and Pearce 2013) are used to select high-order level factors that 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 absolute loadings greater than 0.7. Table 3 illuminates the latent variables derived 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

Review valence, review positivity, and its difference from business/reviewer average rating (positivity) correlate highly and reflect altogether whether a review is in favor of the reviewed business vis-à-vis other reviews, which we name *review valence frame*. The higher rating a reviewer gives, the more likely that rating outstrips business or reviewer average rating, implying his higher favor toward the business compared to an average reviewer and vice versa. Indeed, review rating and sentiment scores are positively correlated (Mankad et al. 2016).

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 User about Reviewer's expertise and the likely helpfulness of the focal review. More precisely, 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), Reviewer 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's number of followers correlates with these reviewer expertise elements in our data. Hence, the latent attribute underlying these variables can be interpreted as *reviewer expertise*.

Neirotti et al. (2016) find that, when User and Reviewer share similar interests or know each other, she tends to trust his review and may like, comment on, or re-read it. On our study's platform, these attributes can be captured by prior reviewer-user interactions such as comments or likes for previous reviews between User and Reviewer. However, our results show that common followship features, measured by the common followers and followees of Reviewer and User, are also captured by the same latent variable as other attributes of reviewer-user interactions. Since these common followship measures can be used to reflect the similarity between a focal user and a potential followee in followee-recommendation literature (Xu et al. 2015), the latent variable capturing both reviewer-user interactions and common followship can be interpreted as *reviewer-user 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) as it allows testing theoretical models for predictive purposes, relaxing normality assumptions, and leveraging latent scores for subsequent analyses (Hair et al. 2019). We carry out convergent validity analysis (Sethi and King 1994) and discriminant validity analysis (Fornell and Larcker 1981), which are commonly adopted (Henseler et al. 2016, Hair et al. 2019) to substantiate scale validity. As can be seen in Table 3, all standardized path coefficients are of acceptable magnitude and statistically significant, implying good convergent validity (Sethi and King 1994, Hair et al. 2019). As regards the discriminant validity, Table 3 illustrates that all factors have good composite reliability, which is above the 0.7 threshold (Fornell and Larcker 1981, Hair et al. 2020). Also, each factor's average variance extracted (AVE) 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 considering only factors in EFA with Cronbach's alpha greater than 0.7, which is considered a lower bound to internal consistency (Sijtsma 2009, Henseler et al. 2016), we believe that the factors reported herein are properly measured by their items, whose contents are relevant to the target latent variables. 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

Overall, the tests above corroborate the validity of the factors arising from our factor analysis. From the original set of variables, we develop eight composite scores and remove those with contents related to the composite scores created. The composite scores include *review valence frame*, *review quality*, *review votes/likes*, *reviewer expertise*, *reviewer social connectedness*, *reviewer-user similarity*, *user following reviewer recently*, and *reviewer following user recently*. Also, there are nine variables, i.e., *brand strength*, *review variance*, *review length*, *review picture*, *review age*, *reviewer locality*, *reviewer-user common locality*, *user dislikes for reviewer*, and *reviewer dislikes for user*, which 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 their correlations in Table 4.

When combining the items 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). So, we proceed with the unweighted average to build our predictive model, which is called *model with unweighted scales* as it is convenient to create and repeat (Bobko et al. 2007). We also perform robustness checks by comparing this model to the one *with weighted scores* where the weight vector for a composite score's items is computed by PLS-SEM (Venturini and Mehmetoglu 2019). The latter's performance is qualitatively similar to the former's.

4.2 Hypothesis Development and Testing

As we obtain the set of independent variables of interest, we next formally develop hypotheses to test if they affect user affinity statistically. The model conceptualized for the three groups of variables identified in Section 4.1, i.e., reviewer characteristics, review attributes, and product features, is illustrated in Figure 2. As many of the reviewer characteristics obtained from the previous step involve unique interactions with users, they help personalize review recommendation. Thus, we formulate hypotheses for them in Section 4.2.1. Sections 4.2.2–4.2.3 elaborate on other variables as control variables and section 4.2.4 presents the hypothesis test results.

Figure 2. Conceptual Framework

4.2.1 Reviewer characteristics

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

The review platform under analysis has a reviewer rank index, but we cannot backtrack its value to the time each review was read, so we do not include it in the model. In our work, reviewer expertise is measured by Reviewer's prior total votes, likes, photos, and followers (dimensionality reduction results). Since reviewer expertise increases review helpfulness, User is more likely to cast a helpfulness vote, heralding her increased affinity for the platform.

H1. Reviewer expertise increases user-review affinity.

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

In addition to expertise, reviewer social connectedness (or social network) has direct and moderating effects on review helpfulness (Zhou and Guo 2017). Social connectedness is defined as the relationships with other platform users and measured by the number of friends on Yelp in Zhou and Guo's (2017) paper. Zhang and Lin (2018) use both the number of friends and followers (fans) on Yelp to operationalize reviewer social networks. On our review platform, this concept can be computed by the *number of followers* and the *number of followees*. Hong et al. (2017) ascertain that the *number of followers* and the *number of followees* have a consistently positive influence on review helpfulness. Aghakhani et al. (2021) log-transform these figures in their model, but their 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 followees of user i on the platform and $fr(u_i)$ denote the set of followers of 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 the network effect posits, reviewers who have many followings are likely to write quality and relevant reviews, which can make readers cast helpfulness votes, pass comments, or re-read, so our hypothesis is:

H2. Reviewer social connectedness increases user-review affinity.

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

Even if User and Reviewer have not established a follower-followee relationship, we can identify prospective followees predicated on followee-recommendation scholarship. Since the followees recommended may well share common interests with User (Armentano et al. 2013, Li et al. 2016), their reviews can be relevant and helpful to her. Since the candidate followees are not yet in User's network, suggesting their reviews to her can boost review recommendation diversity. To identify relevant followees, we can compute the similarity between the focal user (user i) and another user (user $j, j \neq i$) by: $sameFe = \left|fe(u_i) \cap fe(u_j)\right|$; $sameFr = \left|fr(u_i) \cap fr(u_j)\right|$; $indirectFollowship1 = \left|fe(u_i) \cap fr(u_j)\right|$; and $indirectFollowship2 = \left|fr(u_i) \cap fe(u_j)\right|$ (Xu et al. 2015).

By computing these indicators from followee-recommendation studies, we can ascertain the common followship level between User and Reviewer, thereby predicting review relevance. The network effect and collectivism can justify this choice as User is apt to find shared values with her followers or followees, who then have commonality with Reviewer. Given high collectivism in Southeast Asia, where members hold shared values within their group (Hofstede 2001), a review written by a friend can be deemed helpful/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 like, comment on, or re-read his review. Incorporating this variable in our model can help account for possible autocorrelation, where User would continue to like and comment on Reviewer's reviews as she did in prior observations. Further, frequent interactions between users of similar interest clearly boosts their positive attitude to the platform. Therefore, we hypothesize:

H3. Reviewer-user similarity increases user-review affinity.

Yang et al. (2017) find that reviews written by local reviewers, who reside in the vicinity of the reviewed business, are perceived as more helpful. Explanations comprise 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 considered helpful. Additionally, as reviewer region ID is observable on the platform, User can check if Reviewer is from her locality. Given high collectivism in Southeast Asia, where members hold shared values within their group (Hofstede 2001), we can hypothesize that if User and Reviewer have the same region ID, his review is more probably relevant to her, prompting her to cast helpfulness votes, pass comments, or re-read.

H4. Reviewer locality increases user-review affinity.

4.2.2 Review characteristics as control variables

The first control variable of interest is the review valence frame (i.e., the polarity of the review), which has been widely studied. In particular, Quaschning et al. (2015) find in their field and experimental data that the valence of a review significantly affects its helpfulness when it accords with other reviews. In the same vein, Lee et al. (2017) expand on the influence of review valence on review helpfulness and show that reviews with negative valence are usually perceived as more helpful than those with positive valence, but their helpfulness declines when the negative emotions therein are intense. Purnawirawan et al. (2015) ascertain in their meta-analysis that review valence has a significant effect on review helpfulness votes for experience goods and unfamiliar brands.

Next, we investigate the influence of the consistency of review valence (i.e., the variance of the reviews) on user-review affinity. Yelp's strong review helpfulness is indeed ascribable to the high variance in its review sentiment (Xiang et al. 2017). Meanwhile, signaling theory postulates that if a review's valence is widely dispersed from the average rating, that inconsistency might 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 helpfulness votes when there is consistency between the focal review's valence and other reviews'. The aforementioned empirical evidence suggests that high review variance negatively affects review helpfulness perceived by users.

Many scholars (Hu et al. 2017, Lu et al. 2018, Liang et al. 2019) find a direct and significant influence of review quality on review helpfulness, whereas Lee et al. (2018) show that review quality has poor predictive performance for review helpfulness. We measure review quality by the number of votes a review had already received prior to being read by User, which is supported in the literature (Yang et al. 2017). On our work's platform, review votes include likes and 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 that respectively equal the total number of review votes and likes divided by the time lapse in days since review post. As old reviews have more time to accumulate votes, we use the average number of votes and likes to proxy review quality and penalize less recent reviews (Hu et al. 2017, Lee et al. 2018, Tsai et al. 2020).

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 likely conforms with the majority and interacts with such reviews by liking, commenting, or re-reading within seven days.

Another commonly discussed review feature is review age, which is measured in days elapsed since review post (Hu and Chen 2016, Hong et al. 2017, Hu and Yang 2021) and can be rescaled logarithmically (Gao et al. 2017, Aghakhani et al. 2021). While Gao et al. (2017), Hong et al. (2017), and Hu and Chen (2016) find that review age raises the perceived review helpfulness, Wu's (2017) results are mixed, differing by product type, but the aggregate effect is negative. Meanwhile, Yang et al. (2019) illustrate a negative influence, which means older reviews are deemed less helpful. As the businesses reviewed on the platform in question provide hedonic/experience products, our study is in favor of the findings of Wu (2017) and Yang et al. (2019), where review age lowers the helpfulness or relevance of the review for such items and users are less likely to cast helpfulness votes or spend time re-reading or commenting on a less relevant/helpful review.

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. Utilizing this feature in our model also helps us with counterfactual analysis given that review arrangement in the system can be manipulated.

In a separate research vein, Fang et al. (2016), Gao et al. (2017), Hu and Yang (2021), Quaschning et al. (2015), and Wu (2017) illustrate that review length positively affects review helpfulness. 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 as lengthy reviews are less likely to be perused and thus less likely to be assessed. The review length which is readily measured on our work's platform is log-rescaled in our model as in the articles of Aghakhani et al. (2021), Gao et al. (2017), and Karimi and Wang (2017). By including this 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 review groups in our database is *review picture*. According to Ma et al. (2018), Yang et al. (2017), and Zhou and Guo (2017),when a review is accompanied by at least one photo related to the reviewed item, users are more likely to perceive that review as helpful. Filieri et al. (2018) find that the perceived helpfulness of extreme reviews increases when they are long and posted with pictures, which are deemed more convincing than words (Fang et al. 2016).

4.2.3 Product characteristics as a control variable

As can be seen from our discussion in Section 4.2.2, *product type* plays the moderating role (Mudambi and Schuff 2010, Wu 2017). Nonetheless, as all the businesses reviewed on the platform deliver hedonic/experience products in hospitality, product type cannot differentiate review helpfulness in our model, thus not considered. Nevertheless, we can see in the cited literature some other less commonly controlled yet relevant features such as brand similarity (Purnawirawan et al. 2015), hotel features (Anderson and Lawrence 2014, Liang et al. 2019, Filieri et al. 2021), total reviews received (Lee and Choeh 2016, Filieri et al. 2021), product awareness, quality, and popularity (Zhang and Lin 2018), and average rating (Filieri et al. 2021). We find these variables appertain to *brand strength*. Prior research (Sridhar and Srinivasan 2012, Choi and Mattila 2018) has shown that 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 well follow that norm when rating. Another explanation relates to Tsao et al.'s (2019) findings that, for strong brands, negative reviews exert a stronger impact of on sales than positive ones and that management is advised to address such negative reviews. Therefore, the overall ratings are higher for stronger brands. We thus operationalize brand strength by business average rating. There are mixed results for this variable in prior works. Purnawirawan et al. (2015) find that reviews for unfamiliar brands are deemed more helpful while brand strength indices studied by Filieri et al. (2021), Lee and Choeh (2016), and Zhang and Lin (2018) positively influence or moderate review helpfulness, notably for experience goods, which are pertinent to the review platform under analysis in this article. 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.

We use the first six months of the data to test the conceptual model (Figure 2). The last six months' data will be utilized 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 shows the test results for the conceptual model, based on which vital features are input into predictive and counterfactual analyses. We compute the variance inflation factor (VIF) to check if multicollinearity persists after dimensionality reduction. As the VIF is less than 10, the structure of our model is supported (Marquaridt 1970).

Table 5. PLS-SEM Path analysis

Turning first to the group of review features, Table 5 shows that review valence frame creates a significant and positive impact on user-review affinity. In other words, positive reviews are more likely to boost user affinity for the platform. Meanwhile, review variance negatively impacts user-review affinity as expected. Although 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 use either total review votes/likes or average review votes/likes per day to measure review quality. Our paper results support the latter, which is in line with Hu et al. (2017), Lee et al. (2018), and Tsai et al. (2020). Meanwhile, the impact of total review votes/likes is negative, which can be explained by the fact that old reviews have more time to accumulate votes but are considered less helpful. The negative effect of total review votes/likes on user-review affinity is consistent with the negative impact of review age in our data.

With regard to other control variables, review pictures were nonsignificant, whereas review length correlates negatively with user-review affinity, signifying that long reviews are less likely to be liked, commented on, or re-read. Our results also show that brand strength reduces user-review affinity. Since users might have already been familiar with the brand (Purnawirawan et al. 2015), reviews might be considered less impactful.

As regards our hypotheses, reviewer expertise negatively influences user-review affinity, leading to H1 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 (online appendix). However, the interaction term between reviewer expertise and review variance makes a positive impact on user-review affinity. This means that reviewer expertise moderates the relationship between review variance and user-review affinity, in line with the findings of Filieri et al. (2019) and Quaschning et al. (2015), where reviews written by expert reviewers are deemed more helpful when deviating more from business average ratings (see the online appendix).

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

The positive effect of reviewer-user similarity on user-review affinity in Table 5 substantiates H3. This accords with Neirotti et al.'s (2016) findings that when User and Reviewer know each other or find common interests via prior interactions or common followships, she tends to trust his review and likes/comment on/re-read it.

In line with Yang et al. (2017), the positive coefficient of reviewer locality denotes that reviews written by local reviewers are deemed to boost user-review affinity, corroborating H4. Our data also show that user-review affinity increases when User and Reviewer are from the same region, which was understudied in the literature but can be explained by Southeast Asia's high collectivism, where in-group members hold shared values (Hofstede 2001).

Given the tested conceptual model, we replicate Venturini and Mehmetoglu's (2019) PLS-SEM algorithm in Python to calculate the composite scores in Figure 2 as inputs for our predictive and counterfactual analyses.

5. PREDICTIVE MODEL

Given the insights from the previous section, we leverage them for predictive modeling and demonstrate that our results also apply to out-of-sample instances. We begin by considering ML algorithms for our predictive model.

In 18.20% of the instances in our sample, users either liked, commented on, or re-read the review within seven days, so our dataset might epitomize the imbalanced class problem. Based on Napierala and Stefanowski's (2016) categorization, 72.92% of these minority class data points are either "*safe*" or "*borderline*," which can be classified by their nearest neighbors, suggesting that class imbalance may not pose problems for our predictive modeling.

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 its generalization error is theoretically proven in Breiman's (2001) seminal paper and its consistency is substantiated in several recent papers with theoretical analyses (Scornet et al. 2015, Wager and Athey 2018) and empirical findings (Calderoni et al. 2015, Mercadier and Lardy 2019). Scholars report Random Forest's superior performance compared with other methods, such as support vector machine and regression tree (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). Yet, to select a robust model, we compare RFC with some other common algorithms (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 ANN's performance (Lolli et al. 2017, Wang et al. 2018b, Huber et al. 2019). For example, in Arcos-García et al.'s (2017) research, their ANN model performance was not compromised by data imbalance while Huber et al.'s (2019) ANN algorithm can perform well in the presence of relaxed normality assumption provided that the dataset is big enough. Nonetheless, ANNs are often considered a black box (Chen and Hao 2017) with many hyperparameters, e.g., number of neurons and layers, to fine-tune (Ahmad et al. 2017). In Wang et al.'s (2018b) review, 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 build on randomly bootstrapped copies of the original instances, where features for node splitting can be drawn with or without replacement (Louppe and Geurts 2012). Given this added randomness, the correlation between decision trees in the forest decreases and the model performance improves, along with overfitting avoided and variance reduced (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 technique to handle outliers and heterogenous attributes in multidimensional data (de Santis et al. 2017). The algorithm utilizes gradient-based approximations to split the tree node on the negative gradient for loss minimization (Athey et al. 2019), thereby allowing optimizing 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 (2018) is among the earliest papers applying GBC for review helpfulness prediction built on reviewer characteristics and review content variables. Their results show that GBC has lower (root) mean squared error than RFC and ANN.

As regards some hyperparameters selected for our models, which are run on scikit-learn ML 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 example, the optimal number of nodes per tree is 5 (Tsagkrasoulis and Montana 2018), 8 (Zhou and Qiu 2018), 15 (Genuer et al. 2017, Mercadier and Lardy 2019), and 20 (Chen et al. 2018). Ahmad et al. (2017) find that RFC's performance deteriorates after the max 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 (30, 50, 100, and 200) to select the best alternative. To ensure fair comparisons, these hyperparameters are also applied to the BC, GBC, and ANN models where suitable. In particular, the ANN model has three hidden layer (M, 50, 15) for M equal to the number of decision trees.

We iteratively select one month from July to December as a test dataset and bootstrap data from one up to six months before the test set to train the model. The bootstrap data have the same size as the original training data. This bootstrap-train-test procedure is repeated 30 times for each model. Based on the features selected in the prior section, our predictive models are to predict if User will like, comment on, or re-read the review within seven days of the first read. Of particular note is that there is no single model that outperformed others in all three criteria (precision, recall, and F1). The ANN method was the least stable with very large standard deviations compared to other models. BC, GBC, and RFC had similar performance, but GBC's runtime was far longer. Thus, we focus on discussing the BC and RFC results. The complete results 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 reduced dimensionality. The only exception with respect to computational time is RFC, where the processing time difference was only a few minutes. These results imply that our low-dimensional models save substantial runtime with marginal predictive power loss.

We also 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 more data. This suggests that we focus on a smaller yet more recent dataset to save the training time without compromising the predictive model performance. This lends empirical support to 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)

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

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

The confusion matrices in Tables 6–8 present the prediction results averaged over the latter half of year 2017 in our data. The BC models made more positive predictions, whereas their RFC counterparts made overall more true-positive (TP) predictions and less false-positive predictions, leading to a higher precision. The F1 rates and forecast accuracy (TP + true negatives) of the RFC models were also higher. This suggests 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 its recommendation diversity. We will use RFC and BC for counterfactual analysis of our personalized review recommendation.

6. COUNTERFACTUAL ANALYSIS

Based on our exploratory analysis, we propose that reviews be recommend based on their propensity to be engaged by platform users. In the preceding section, we demonstrate that our predictive models (BC and RFC) can predict user affinity for each review consistently and can be used to run counterfactual what-if analysis (Dickerman and Hernán 2020). In particular, for each unique platform user, reviews triggering higher estimated user-review affinity based on the conceptual model parameters are put before those with lower user-review affinity. This *affinity-based ranking* will replace the *original* 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. This counterfactual what-if simulation is to illuminate the performance of our personalized review commendation system.

Reviews with ranking from 1st to 10th (first-page reviews) are considered promoted in our analysis. Because the platform can change this arrangement, we want to test if user-review affinity will grow 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 each user concerned. Based on the confirmed conceptual model parameters (Table 5) and a data subset, 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 in descending order. Since reviewer-user similarity, shared locality, and prior interactions (following and dislikes) vary by reviewer-user pair, each platform user would see a different set of promoted reviews. Given our model's particular relevance for businesses with so many reviews that users may face information overload, our data subset for counterfactual analysis focuses on those with at least 50 reviews (five review pages).

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

Table 9 shows 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 (*t*-test, KS test, and *z*-test) indicated consistent results: raising or lowering this threshold by 10 reviews did not alter the statistical comparability of those subsets (see Table 9). The counterfactual analysis results reported in Tables 10 and 11 are for businesses which had at least 50 reviews. In our counterfactual analysis, a business is considered to have non-decreased user-review affinity when its positive user interactions (likes, comments on, or re-reading of its reviews) simulated with affinity-based ranking are greater than or equal to those with original ranking.

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

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

As can be seen in Tables 10 and 11, the reviews promoted by the novel affinity-based ranking, which builds on the conceptual model, increased user affinity to the platform (by liking, 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, for businesses with at least 50 reviews, the re-ranking increased user interactions in all simulators. Users reading the promoted reviews for those businesses also interacted more with those reviews and that jump in interactions was statistically significant at the 1% level in most simulators. Thus, review platforms can leverage this insight to rearrange product reviews in a personalized fashion for each user to boost user-review affinity. Moreover, in line with Tables 6 and 7, the BC-based system produced more recommendations and thus likely boosted the diversity of recommended reviews.

7. DISCUSSIONS AND CONCLUSIONS

Online reviews have become an integral part of many online platforms. While spending considerable resources attracting users to contribute online reviews, these platforms encounter critical issues where their customers have too many reviews to read, leading them to suffer from information fatigue. We propose alleviating such issues by developing a personalized review recommendation framework that can help platforms selectively display reviews to their user based on the propensity that the user will engage with each review.

We begin by conducting an exploratory analysis where we survey previous research to identify key independent variables that can affect user affinity (i.e., the tendency that a user would like, comment on, or re-read the reviews). To reduce high dimensionality and avoid multicollinearity, we conduct factor analysis and confirmatory composite analysis. Based on the results, we corroborate several important features, notably reviewer-user similarity, (shared) locality, and followship, which are crucial yet underexplored in the review-recommendation literature.

Following that, we leverage the insights uncovered from our exploratory analysis for predictive modeling. Here, our goal is to ensure that our insights apply to out-of-sample instances to verify the external validity of our findings. In addition, this exercise allows us to predict the propensity that each user would interact with each review, which is the key ingredient used for our personalized review recommendation system in the next step. With a consistently accurate predictive model, we proceed to counterfactual analysis where we re-rank reviews based on their potential user affinity. Our counterfactual simulation results illustrate that re-ranking reviews can attain significantly more user engagement, which generally leads 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.
Anderson CK, Lawrence B (2014) The influence of online reputation and product heterogeneity on service firm financial performance. *Service Science* 6(4):217–228.
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, Cambazoglu BB, Gullo F, Mantrach A, Silvestri F (2017) Exploiting search history of users for news personalization. *Information Sciences* 385-386:125–137.
Bansal P, Gualandris J, Kim N (2020) Theorizing Supply Chains with Qualitative Big Data and Topic Modeling. *Journal of Supply Chain Management* 56(2):7–18.
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.
Caro F, Kök AG, Martínez-de-Albéniz V (2020) The future of retail operations. *Manufacturing & Service Operations Management* 22(1):47–58.
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.
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.
Choi HS, Oh W, Kwak C, Lee J, Lee H (2022) Effects of online crowds on self-disclosure behaviors in online reviews: A multidimensional examination. *Journal of Management Information Systems* 39(1):218–246.
Choi S, Mattila AS (2018) The effect of experience congruity on repurchase intention: the moderating role of public commitment. *Service Science* 10(2):124–138.
Chua AYK, Banerjee S (2017) Analyzing review efficacy on Amazon.com: Does the rich grow richer? *Computers in Human Behavior* 75:501–509.
Cohen MC (2018) Big data and service operations. *Production and Operations Management* 27(9):1709–1723.
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.

Farrell J, Klemperer P (2007) Coordination and lock-in: Competition with switching costs and network effects. Armstrong M, Porter R, eds. *Handbook of Industrial Organization*, vol. 3 (Elsevier), 1967–2072.
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, Ash Ray P (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 M-H, Rust RT (2013) IT-Related Service: A Multidisciplinary Perspective. *Journal of Service Research*

16(3):251–258.
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, Fu WW (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.
Khan I, Rahman Z (2017) Brand experience and emotional attachment in services: The moderating role of gender. *Service Science* 9(1):50–61.
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.
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 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.
Mankad S, Han HS, Goh J, Gavirneni S (2016) Understanding online hotel reviews through automated text analysis. *Service Science* 8(2):124–138.
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) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly* 34(1):185–200.
Napierala K, Stefanowski J (2016) Types of minority class examples and their influence on learning classifiers from imbalanced data. *Journal of Intelligent Information Systems* 46(3):563–597.
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 (2021) Mitigating the adverse effect of monetary incentives on voluntary contributions online. *Journal of Management Information Systems* 38(1):82–107.
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, Janze C (2019) Information processing on online review platforms. *Journal of Management Information Systems* 36(4):1347–1377.
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.
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.
Spence M (1973) Job Market Signaling. *The Quarterly Journal of Economics* 87(3):355–374.
--- (2002) Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review* 92(3):434–459.
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 DJ, Benbasat I, Cenfetelli RT (2017) A two-stage model of generating product advice: proposing and testing the complementarity principle. *Journal of Management Information Systems* 34(3):826–862.
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.
Yayla-Küllü HM, Tansitpong P, Gnanlet A, McDermott CM, Durgee JF (2015) Employees' national culture and service quality: An integrative review. *Service Science* 7(1):11–28.
Yong AG, Pearce S (2013) A beginner's 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 (2022) Social interactions and peer evaluation on online review platforms: An empirical analysis. *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.

Tables

Table 1. Descriptive statistics of the dataset

		Number of interactions recorded				
	Unique number	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

Table 2. List of variables

Group	Definition/Operationalization/Feature	Prior studies
Review features	Review valence: the star ratings of the review (range between 1 and 5).	[16, 18]
	Review positivity: 1 if the rating is greater than 3, -1 if less than 3, 0 otherwise.	[6, 19]
	Difference between review valence and business average rating.	[3, 11, 28]
	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.	[18, 23]
	Review helpfulness score: The (average) number of (helpfulness) votes that the review received.	[10, 22, 29]
	Review age: The time difference between review posting and review reading.	[7, 9, 10]
	Review length: The number of words/characters in the review (measured by the platform in question).	[1, 7, 12]
	Review picture: The number of pictures in the review.	[4, 15, 25]
Reviewer characteristics	Reviewer's total number of prior reviews.	[4, 29]
	Reviewer's total number of reviews with quality flag.	[5]
	Reviewer's total number of photos.	[3, 4]
	Reviewer social connectedness (or reviewer social network).	[9, 29]
	Reviewer's number of followers.	[1, 9, 26]
	Reviewer's number of followees.	
	Reviewer's number of followings or Reviewer 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).	[13, 14, 17, 21]
	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).	[27]
	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.	

Group	Definition/Operationalization/Feature	Prior studies
	Reviewer's comments on User's posts before.	
	Reviewer-User common followers.	[24]
	Reviewer-User common followees.	
	Reviewer-User indirect followships: Reviewer's followers are User's followees and vice versa.	
	Reviewer locality: If the reviewer is a local in the region of the reviewed business, reviewer locality is 1, 0 otherwise.	[25]
Product attributes	Brand strength: The business average rating.	[2, 8, 20]
[1]=(Aghakhani et al. 2021), [2]=(Blal and Sturman 2014), [3]=(Fang et al. 2016), [4]=(Filieri et al. 2018), [5]=(Filieri et al. 2019), [6]=(Filieri et al. 2021), [7]=(Gao et al. 2017), [8]=(Ho-Dac et al. 2013), [9]=(Hong et al. 2017), [10]=(Hu and Chen 2016), [11]=(Hu et al. 2008), [12]=(Karimi and Wang 2017), [13]=(Lee et al. 2015), [14]=(Liu et al. 2019), [15]=(Ma et al. 2018), [16]=(Purnawirawan et al. 2015), [17]=(Qian et al. 2014), [18]=(Quaschning et al. 2015), [19]=(Sparks and Browning 2011), [20]=(Tsao et al. 2019), [21]=(Wang et al. 2018a), [22]=(Wu 2017), [23]=(Xiang et al. 2017), [24]=(Xu et al. 2015), [25]=(Yang et al. 2017), [26]=(Yu et al. 2018), [27]=(Yu et al. 2022), [28]=(Zhang et al. 2013), [29]=(Zhou and Guo 2017)		

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 (rValence)	0.935
	Review positivity ($\mathbb{I}(\text{rValence} > 3) - \mathbb{I}(\text{rValence} < 3)$)	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 Reviewer-User common followers	0.843
	Log of Reviewer-User common followees	0.782
	Log of Reviewer-User indirect followship 1	0.840
	Log of Reviewer-User 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's 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).

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)
.83																
–.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	.87									
–.02	–.03	0.48	0.13	0.09	0.11	–.13	0.15	.77								
0.00	–.01	0.07	0.01	0.01	0.02	–.03	0.01	0.07	.83							
–.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	.85					
–.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	.92			
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) Reviewer-user common locality; (17) Brand strength.

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) Reviewer-user common locality	0.0191	0.000	1.021
(17) Brand strength (busAvgRating)	–0.0029	0.000	1.110
(18) Review variance × Reviewer expertise	0.0117	0.000	1.030

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

Actual / Predicted	Positives			Negative		
Positives	356001.03	351721.60	353416.53	30020.57	30984.17	31036.77

	(0.19%)	(0.21%)	(0.19%)	(0.98%)	(1.03%)	(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)

Predicted \ Actual	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.

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

Indicators \ Businesses with	≥ 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% 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% level.

Figures

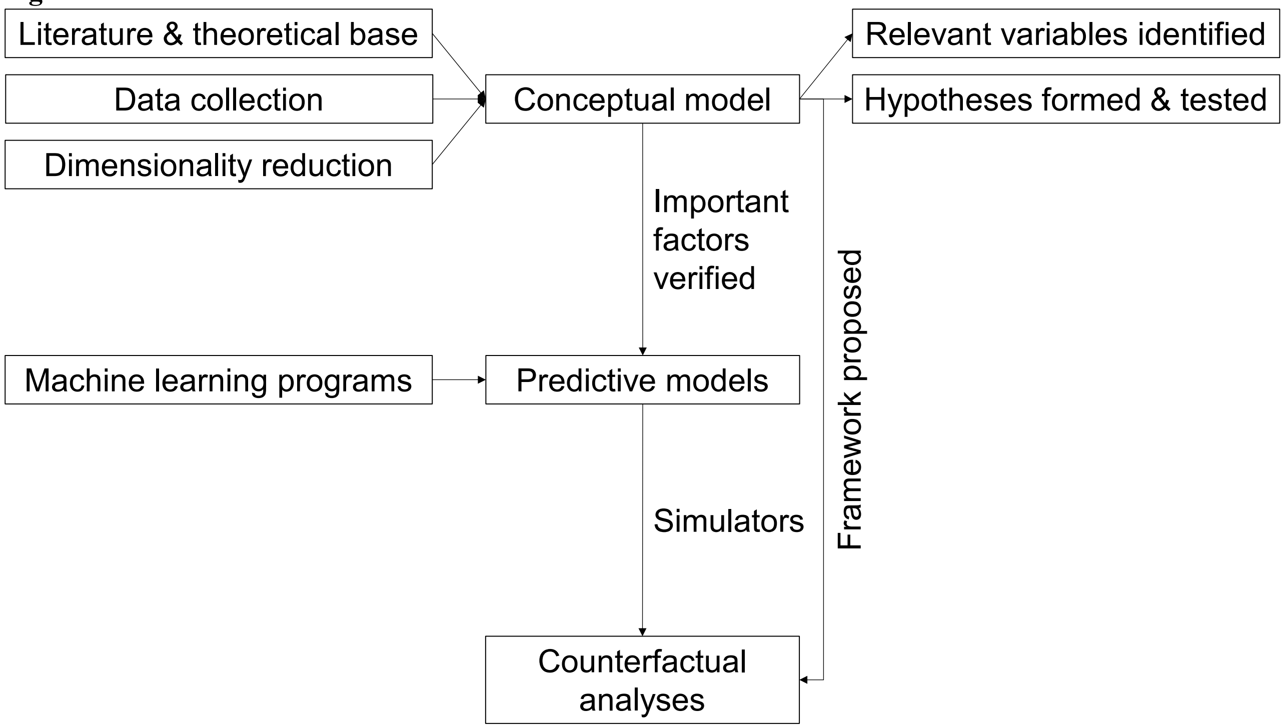


Figure 1. Research design

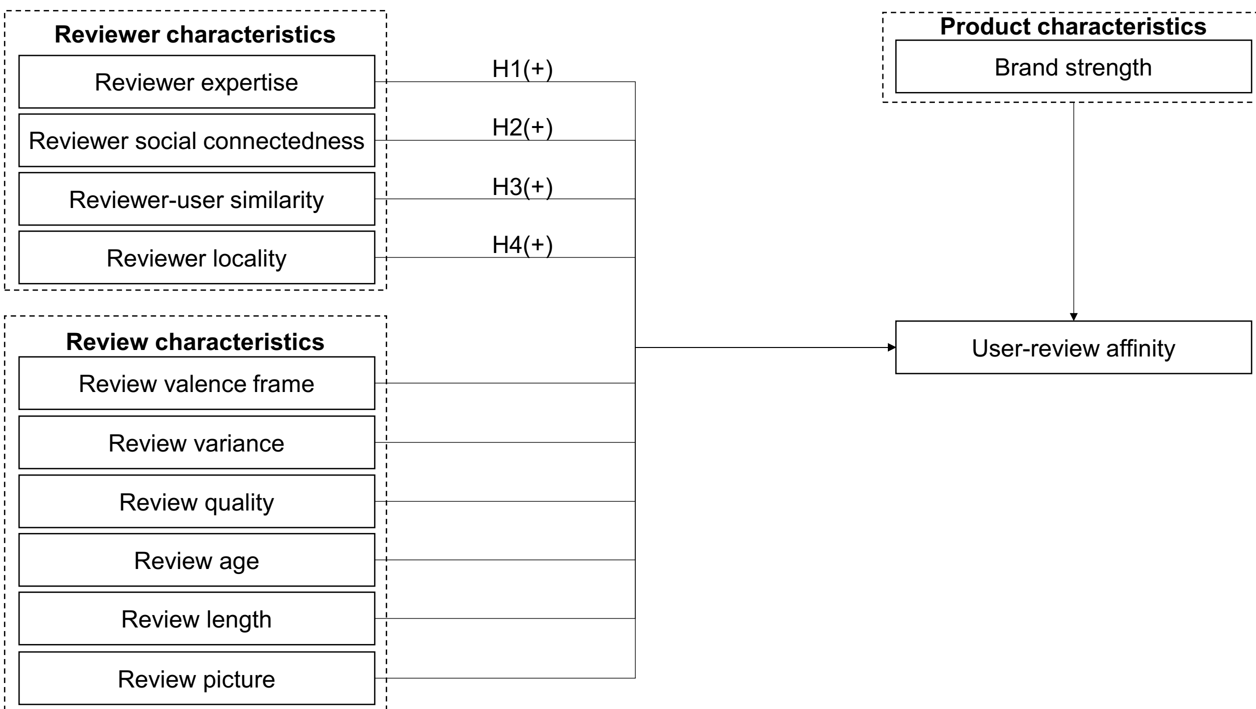


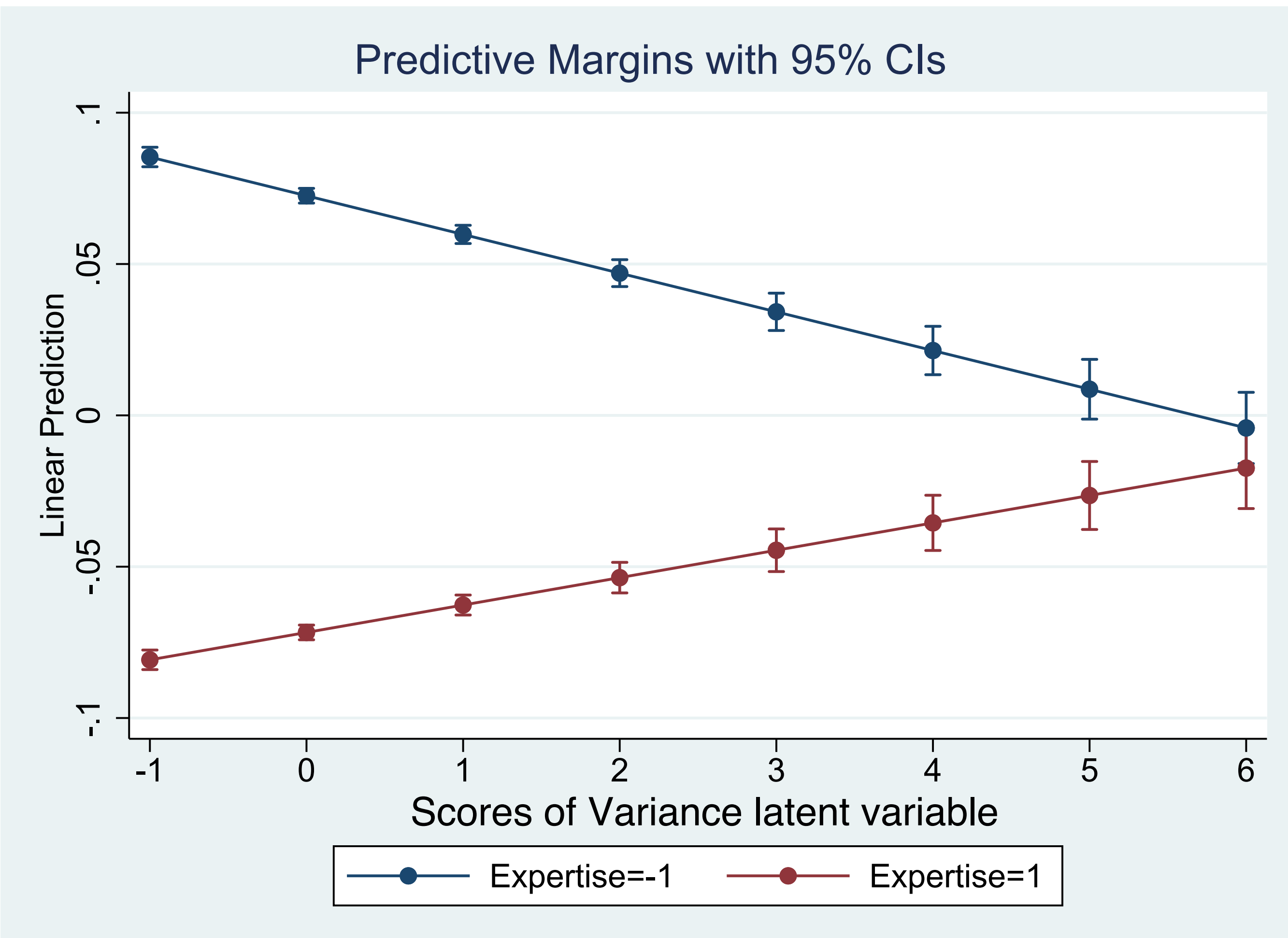
Figure 2. Conceptual Framework

Supplemental material of the paper titled

“**A framework for affinity-based personalized review recommendation**”

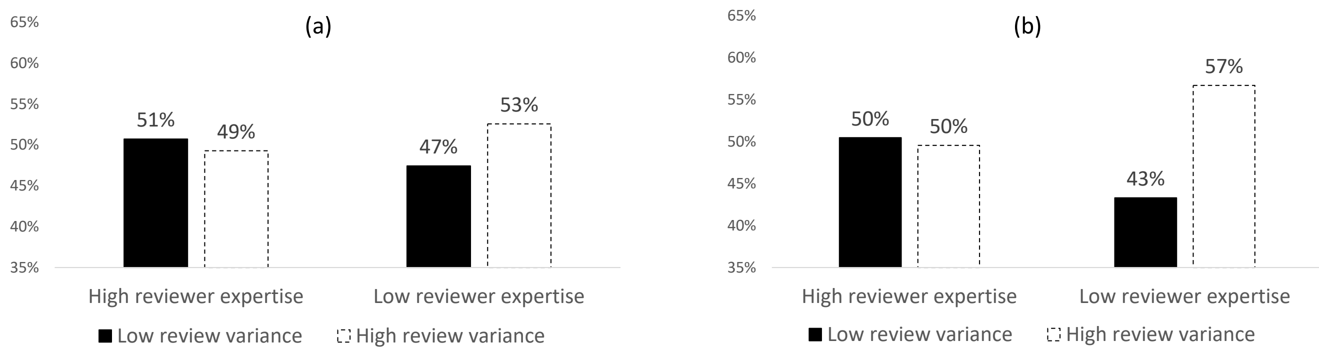
Provided by the authors

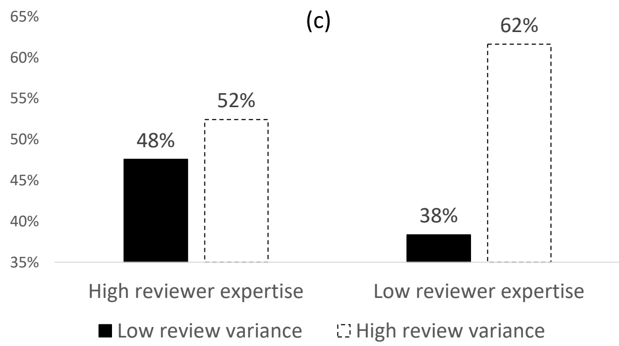
Figure I. Interaction between review variance and reviewer expertise



The negative interaction term between reviewer expertise and review variance means that reviewer expertise moderates the relationship between review variance and user-review affinity, where reviews written by

Figure II. Review variance plotted by reviewer expertise





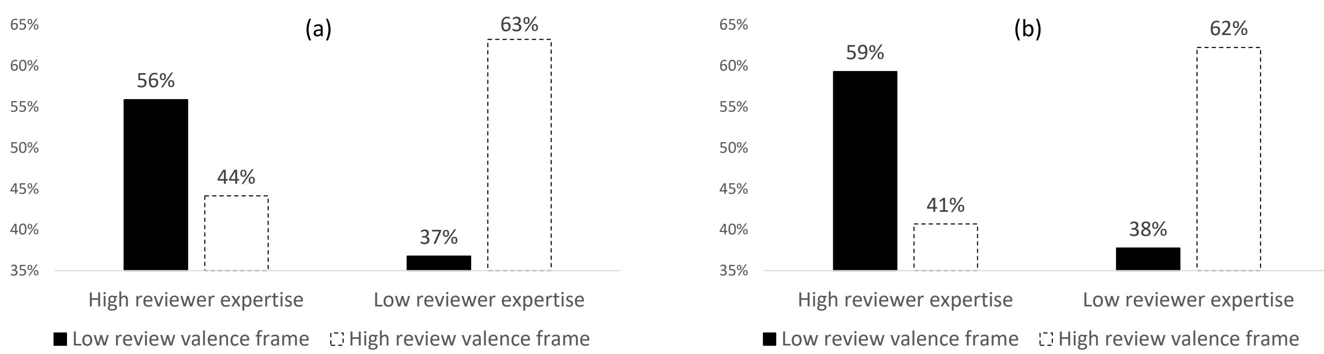
(a) Low = from the 10th percentile to the 25th percentile
High = from the 75th percentile to the 90th percentile

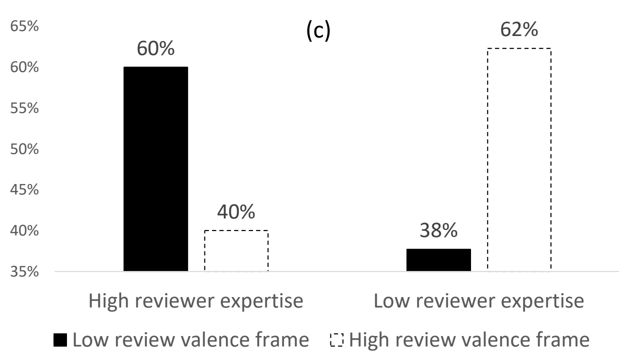
(b) Low = from the 5th percentile to the 25th percentile
High = from the 75th percentile to the 95th percentile

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

expert reviewers are deemed more helpful when deviating more from business average ratings (see Figure I where [reviewer] Expertise and [review] Variance are standardized). In our dataset, reviews written

Figure III. 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

by reviewers with high expertise (reviewer expertise from the 75th percentile to the 90th, 95th, or 99th percentile) did not usually deviate largely from business average ratings compared to their counterparts written by reviewers with low expertise (reviewer expertise from the 1st, 5th, or 10th percentile to the 25th percentile) as depicted in Figure II. Nevertheless, 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 non-expert counterparts (see Figure III) while high review valence frame increases user-review affinity in our results. This might explain why reviewer expertise correlated negatively with user-review affinity in our data.

The evaluation of the machine learning algorithms used in the manuscript, averaged over 30 runs for the dataset from July to December 2017 trained on 1–6 months before, is depicted in the following figures:

- Bagging Classifier (BC): Figures IVa–IVf;
- Gradient Boosting Classifier (GBC): Figures Va–Vf;
- Multi-layer Perceptron Classifier (ANN): Figures VIa–VIf;
- Random Forest Classifier (RFC): Figures VIIa–VIIf.

The additional path analyses mentioned in the manuscript are provided in Table I.

Table I. Additional path analyses

Dependent variable = User-review affinity	Model 1A		Model 1B		Model 2A		Model 2B	
Number of observations	1813477		1813477		1813477		1813477	
Average R-squared	0.04550		0.04551		0.04525		0.04527	
Average communality	0.88067		0.87978		0.87340		0.87252	
Absolute GOF	0.19667		0.19657		0.19507		0.19498	
Relative GOF	0.91874		0.91919		0.91820		0.91866	
Average redundancy	0.04550		0.04551		0.04525		0.04527	
Variable	Coeff.	VIF	Coeff.	VIF	Coeff.	VIF	Coeff.	VIF
(1) Review valence frame	0.022*	1.359	0.022*	1.359	0.023*	1.354	0.023*	1.354
(2) Review variance	–0.002†	1.290	–0.002†	1.290	–0.002‡	1.287	–0.002‡	1.287
(3) Review quality	0.073*	1.713	0.072*	1.714	0.074*	1.713	0.074*	1.714
(4) Review votes (likes)	–0.011*	2.474	–0.011*	2.474	–0.013*	2.467	–0.013*	2.468
(5) Review length	–0.028*	1.542	–0.028*	1.542	–0.032*	1.508	–0.032*	1.508
(6) Review picture	0.0005	1.508	0.0006	1.508	0.0011	1.517	0.0011	1.517
(7) Review age	–0.004*	1.401	–0.004*	1.401	–0.008*	1.413	–0.008*	1.413
(8) Reviewer expertise	–0.072*	1.926	–0.072*	1.926	–0.061*	1.501	–0.061*	1.501
(9) Reviewer-User similarity	0.161*	1.432	0.161*	1.433	0.160*	1.431	0.161*	1.432
(10) User following Reviewer recently	0.019*	1.052	0.019*	1.052	0.019*	1.052	0.019*	1.052
(11) User disliked Reviewer	–0.008*	1.049	–0.008*	1.049	–0.008*	1.049	–0.008*	1.049
(12) Reviewer following User recently	0.014*	1.060	0.014*	1.060	0.015*	1.060	0.015*	1.060
(13) Reviewer disliked User	–0.006*	1.047	–0.006*	1.047	–0.006*	1.047	–0.006*	1.047
(14) Reviewer social connectedness	0.034*	2.193	0.034*	2.193	0.022*	1.986	0.022*	1.986
(15) Reviewer locality	0.003*	1.007	0.003*	1.007	0.004*	1.007	0.004*	1.007
(16) Reviewer-User common locality	0.019*	1.021	0.019*	1.021	0.020*	1.019	0.020*	1.019
(17) Brand strength (busAvgRating)	–0.003*	1.110	–0.003*	1.110	0.0002	1.098	0.0002	1.098
(18) Review variance × Reviewer expertise	0.012*	1.030	0.012*	1.030	0.010*	1.028	0.010*	1.028

Note: Model 1 uses the total count of (good) reviews, photos, and followers to measure (8). Model 2 uses the average count per day for (8). Model A uses the unweighted scale for user's and reviewer's recent votes for each other in measuring (9). Model B uses the weighted score for (9). *: significance at the 1% level. †: significance at the 5% level. ‡: significance at the 10% level.

Figure IV. Evaluation of Bagging Classifier (BC) trained on 1–6 months before (30 runs)

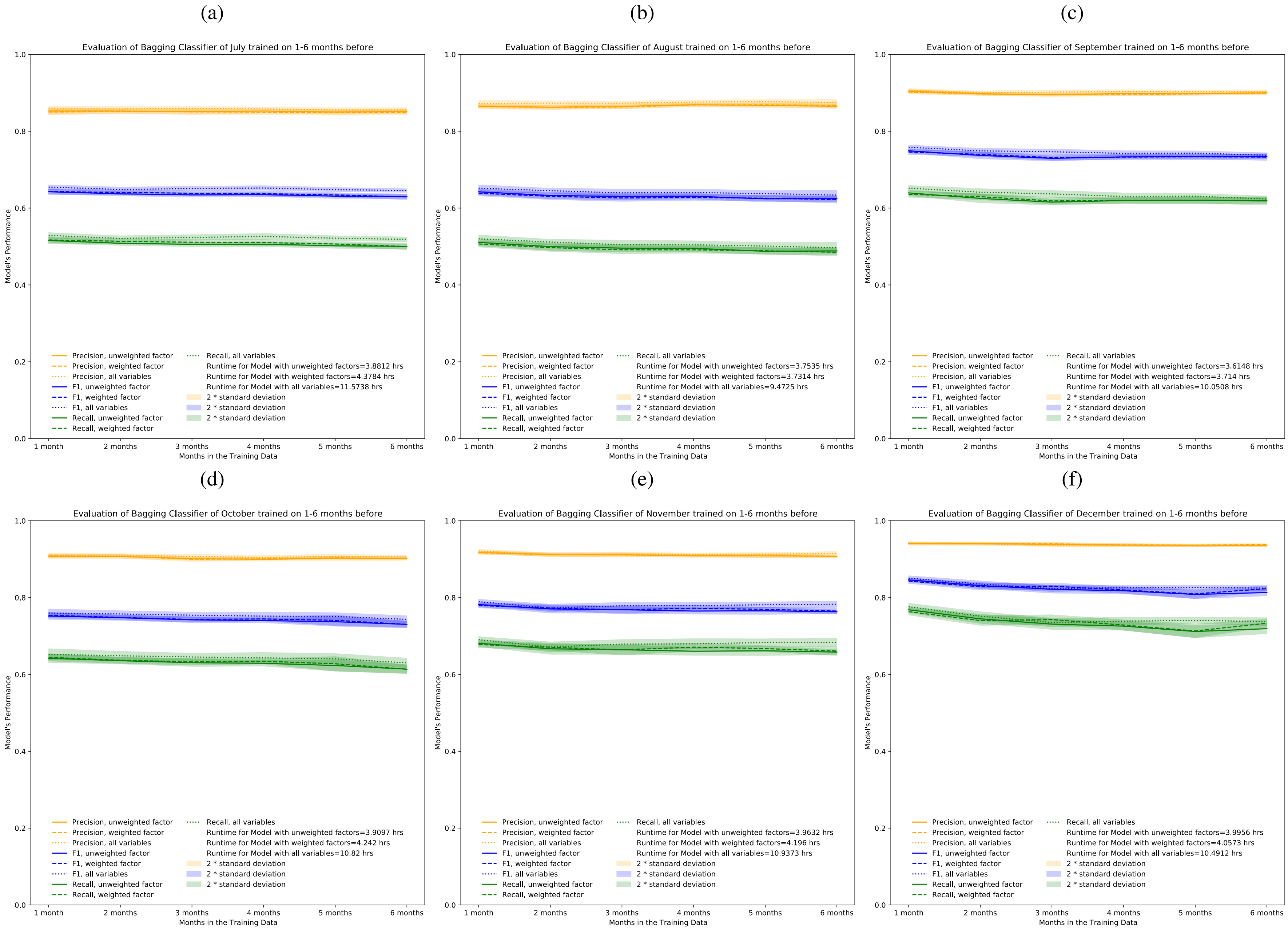
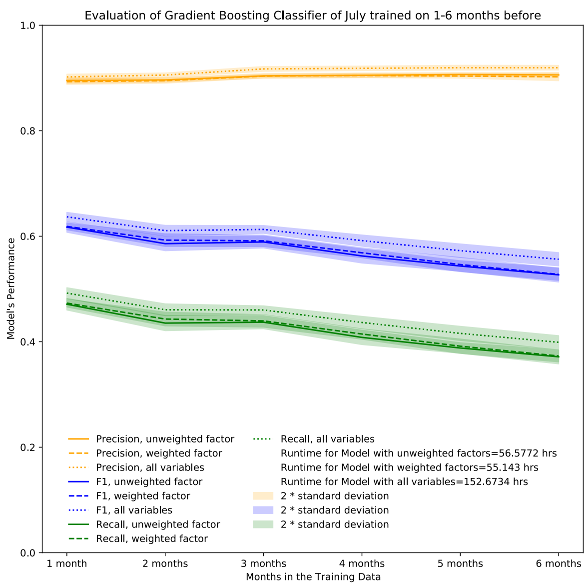
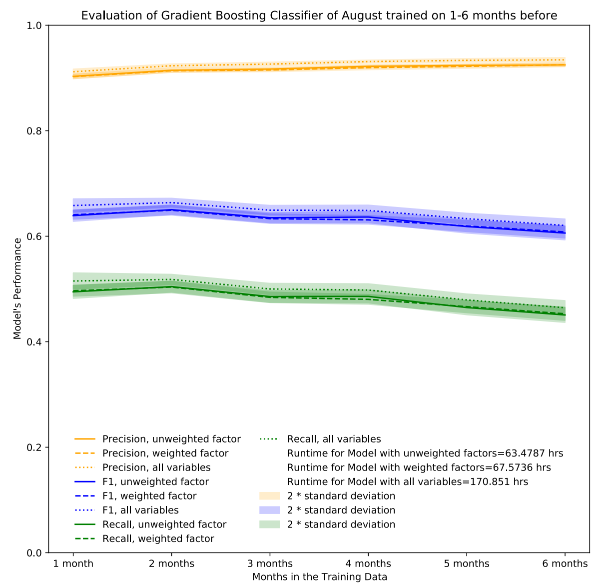


Figure V. Evaluation of Gradient Boosting Classifier (GBC) trained on 1–6 months before (30 runs)

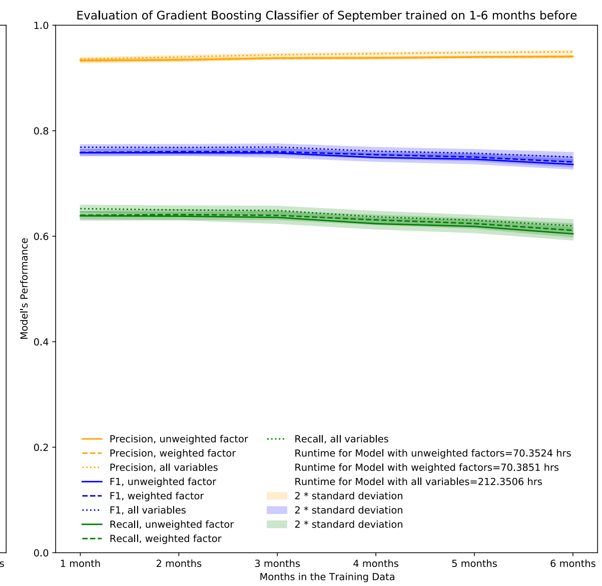
(a)



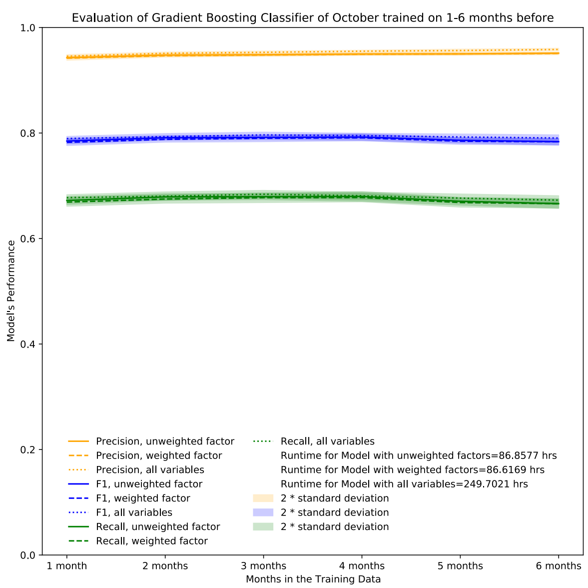
(b)



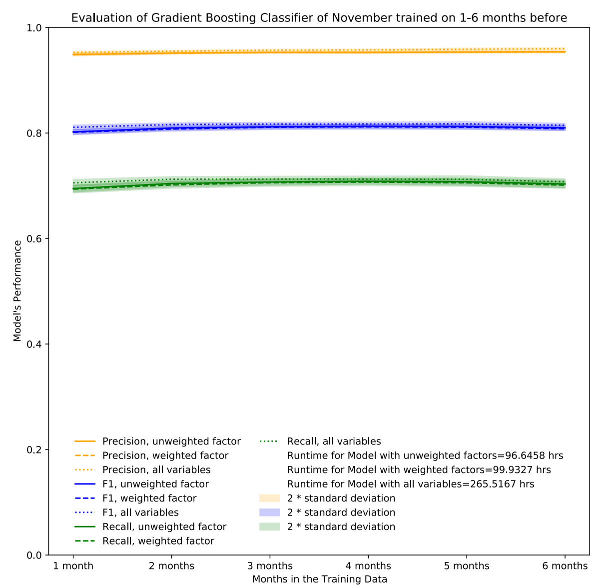
(c)



(d)



(e)



(f)

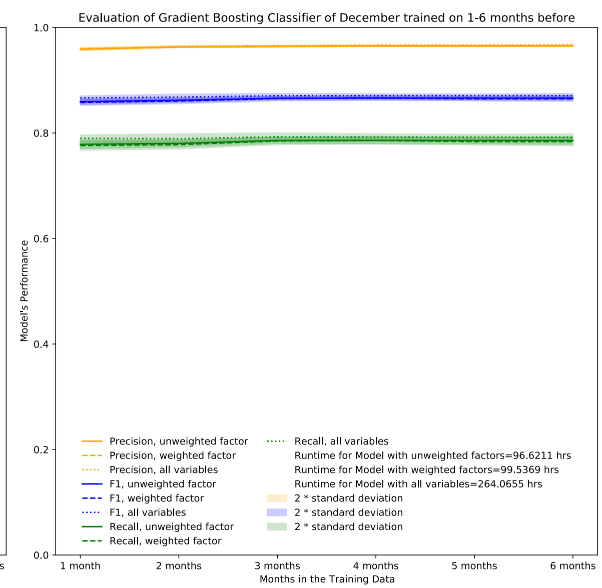


Figure VI. Evaluation of Multi-layer Perceptron Classifier (ANN) trained on 1–6 months before (30 runs)

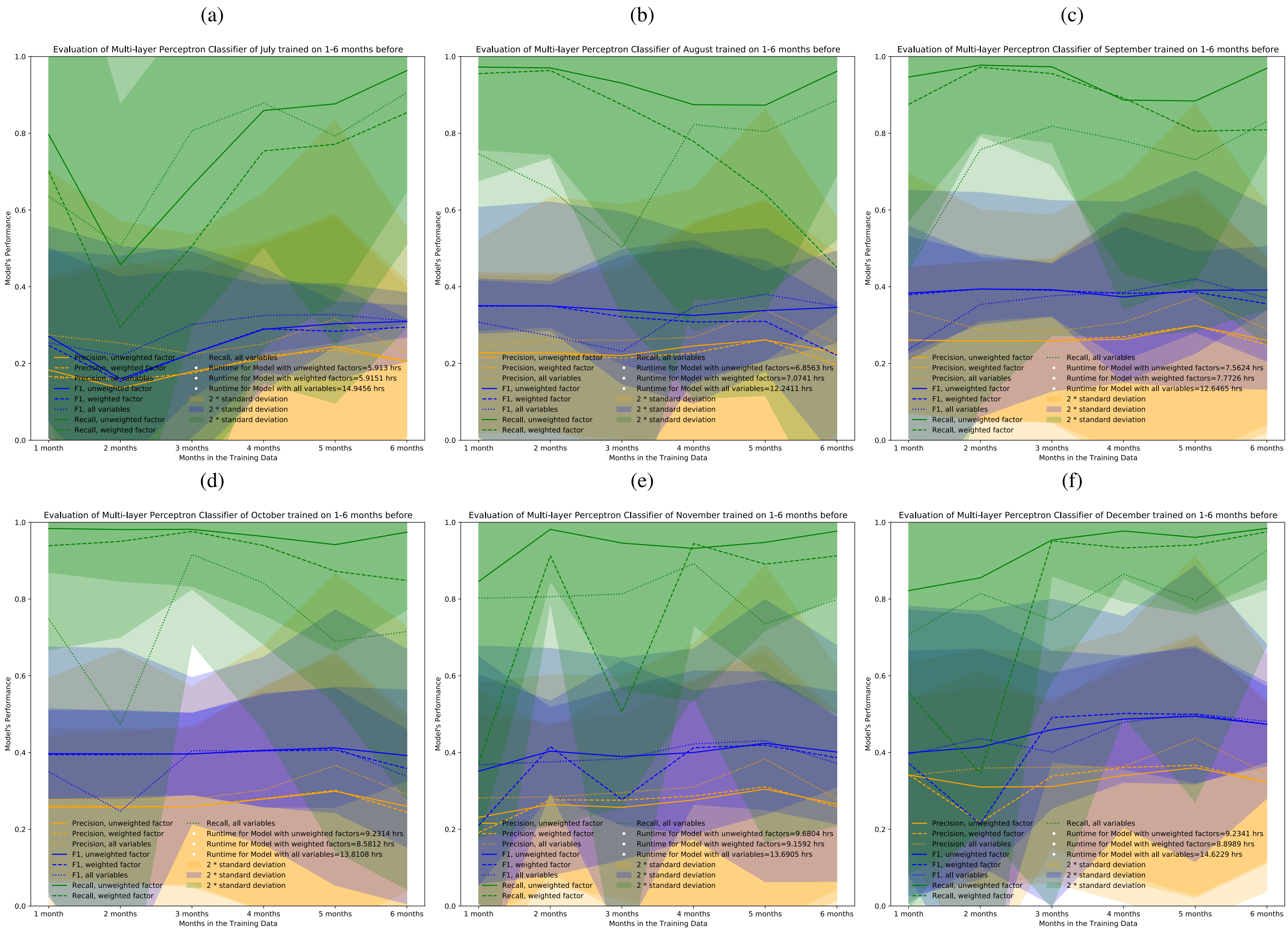
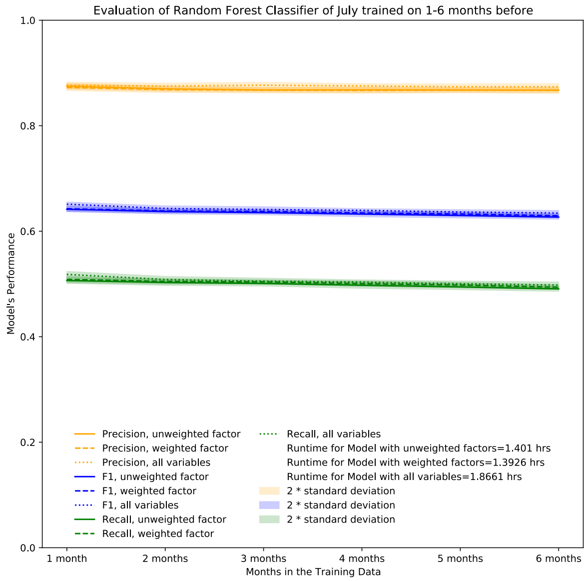
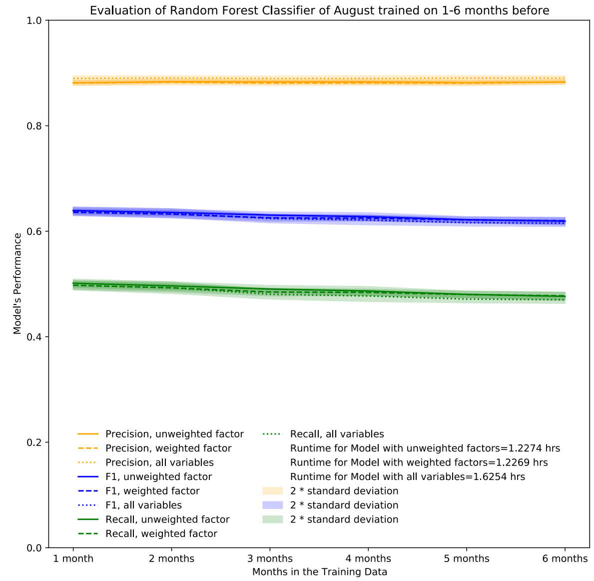


Figure VII. Evaluation of Random Forest Classifier (RFC) trained on 1–6 months before (30 runs)

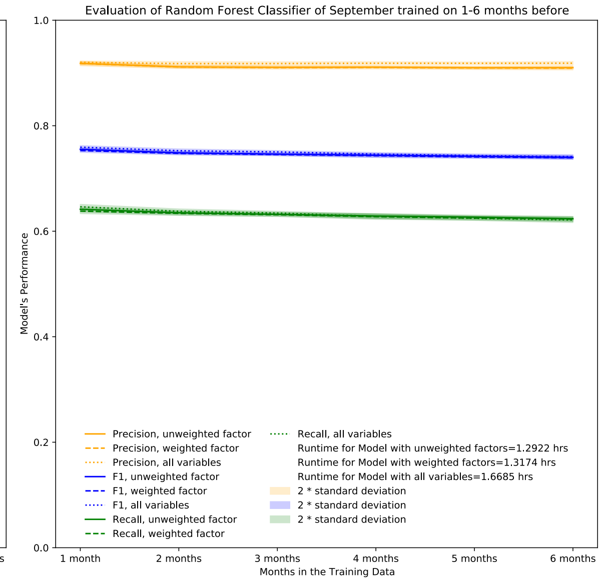
(a)



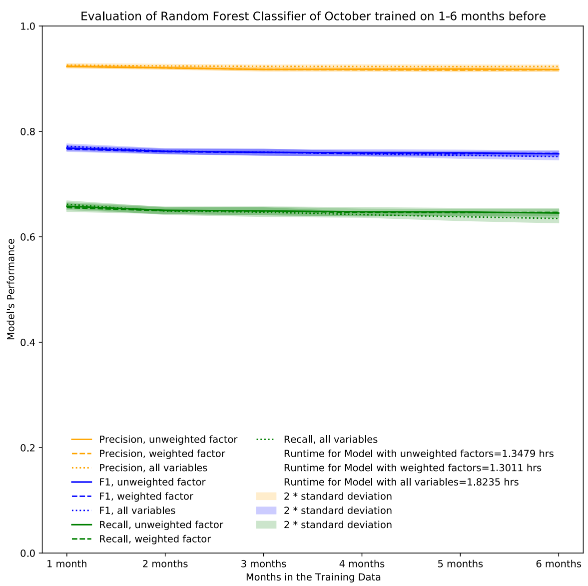
(b)



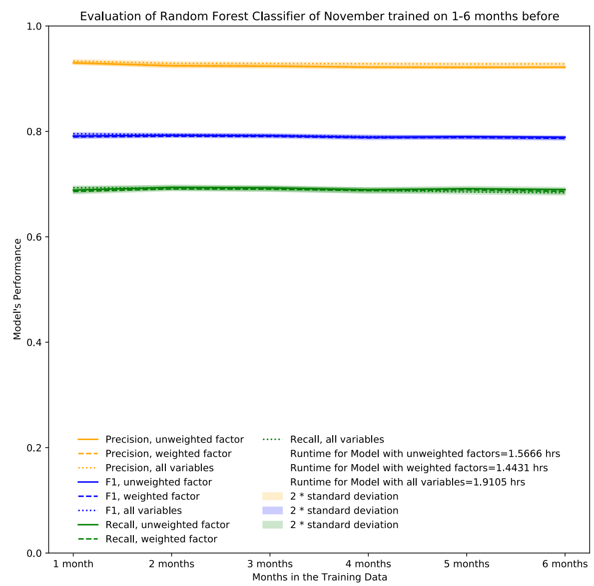
(c)



(d)



(e)



(f)