

## Moderating Effects of Time-Related Factors in Predicting the Helpfulness of Online Reviews: a Deep Learning Approach

Morteza Namvar  
The University of Queensland  
m.namvar@uq.edu.au

James Boyce  
The University of Queensland  
james.boyce@uqconnect.edu.au

Yuanyuan Zheng  
TAO Solutions  
michelle.zheng@taosolutions.ca

Jatin Sarna  
The University of Queensland  
j.sarna@uqconnect.edu.au

Alton Chua Yeow Kuan  
Nanyang Technological University  
altonchua@ntu.edu.sg

Sina Ameli  
Youi Insurance  
sinaa@youi.com

### Abstract

*Given the importance of online reviews, as shown by extensive research, we address the problem of predicting the helpfulness of online product reviews by developing a comprehensive research model guided by the theoretical foundations of signaling and social influence theories. We use review order and time interval to incorporate the moderating effects of the time-related variable on the reviewer's valuation of products and the related details they provide. Applying deep learning techniques in text processing and model building on a dataset of 239297 reviews, the empirical findings represent a strong support of the proposed approach and show its superior performance in predicting review helpfulness compared to current approaches. This research contributes to theory by analyzing online reviews from the points of two well-known information processing theories and contributes to practice by developing a model to sort the newly posted reviews.*

### 1. Introduction

Online reviews play an increasingly important role in developing trust with customers, as they have become useful sources of information and have impacted market transactions and consumption behaviors [1]. An online review typically consists of a star rating, helpfulness votes by users and written comments regarding the product, or even feedback or service experience [2]. Online reviews provide electronic word-of-mouth (eWOM) to reduce the level of perceived uncertainty for products. A customer survey conducted by BrightLocal demonstrated that 82% of participants read online reviews before they make purchase decisions [3]. In addition, online reviews have been found to be significantly related to sales volume and business revenue. Helpful reviews have proved to be more influential on customers' purchase decisions, and thus contribute heavily to a business's profitability [4].

Using a large number of reviews can provide comprehensive information for customers to make a sound decision, however, they mostly overwhelm readers, imposing higher search costs and devaluing the usefulness of reviews [5]. Moreover, conflicting reviews confuse customers to evaluate products appropriately [6]. Thus, a sorting feedback mechanism screening the review helpfulness is needed to assist customers in their purchase decision-making. Sun et al. [7] discussed that review helpfulness could reduce the level of customer uncertainty when they read review contents. They used the ratio of helpful votes to total votes to determine the threshold of classifying helpful and unhelpful reviews. To identify the optimal threshold, Ghose and Ipeirotis [8] conducted experiments and found the optimal ratio of helpful votes to total votes is 0.6. Several other researchers adopted this threshold as well in their studies (e.g., [9, 10]).

Chua and Banerjee [11] proposed that the perception of review helpfulness can be impacted by information quality, which is reflected through review length, review depth, specificity, and reliability. In terms of reviewer-related factors, Hong et al. [2] found that the disclosure of reviewer identity and reviewer expertise positively influence review helpfulness. Ngo-Ye and Sinha [12] suggested not only review text but also reviewer engagement, represented by reviewer activity and commitment, are important factors in determining the helpfulness of reviews.

While there have been several investigations identifying the impact of reviewer-related factors on the helpfulness of review, the role of time-related factors in predicting the helpfulness of online reviews is underexplored. Time-related factors have important implications for our understanding of the phenomenon as the existing studies have conceptualized dynamics as occurring solely as a function of one factor. This study, therefore, draws on signaling and social influence theories to incorporate the impact of time-related factors

and propose a novel approach in predicting the helpfulness of online reviews.

Our research makes several contributions and improvements to the literature of online review helpfulness and practice. We incorporate the impacts of time-related variables in helpfulness prediction and explain them through the lens of two well-known theories. We develop new variables related to reviews using deep learning methods, which have not been discussed in the existing literature of online reviews.

We use a dataset of 239297 reviews from verified purchases to develop a reliable predictive model. The results show the superior performance of our method compared to the existing works on review helpfulness prediction.

Next, we present the background of related studies in the literature, followed by a representation of the details of the research method. Then, we present the empirical results of the study with corresponding discussions. Finally, the paper concludes and elaborates on the limitations of this study and opportunities for further research.

## 2. Background

The extent of the literature of online reviews shows several studies that investigate the helpfulness of online reviews. In terms of review-related factors, Park and Nicolau [1] implemented an empirical application on 5,090 reviews of restaurants and found that reviews with extreme ratings (positive or negative) are considered more helpful than those with moderate ratings. Cao et al. [13] showed similar results by analyzing 3,460 reviews from a software program downloading website. Review length could also impact review helpfulness. Lutz et al. [14] found that longer reviews are more likely to contain frequent changes between positive and negative aspects, which often confuse review readers.

Zhou and Guo [15] identified that the helpfulness of reviews could be declined throughout time. Reviews

sorted as most recent and ranked at the top tend to get more attention, and following reviews then turned out to be less helpful. Through a meta-analysis, Hong et al. [2] showed that review age has a stronger effect on review helpfulness when it is measured by helpfulness vote ratio or when reviews are retrieved from an internal review platform.

Other helpfulness determinants widely discussed by scholars are reviewer engagement, and reviewer reputation. Thakur [16] suggests that higher reviewer engagement will mediate the relationship between customer satisfaction and trust in online reviews. The more reviewers are engaged, the more genuine and trustworthy the user-generated content would be. Thus, customers who read those reviews find them more helpful. In the study by Chua and Banerjee [17], reviewers with a proven track record in writing reviews tend to possess a higher reputation and authority and therefore will attract more trust, as they are believed to be a credible source of information. Lee and Choeh also proposed that reviewer reputation is an important determinant for helpfulness [18].

While the above studies explained the factors that influence the review helpfulness, some other studies have used predictive models and have shown the level of accuracy in predicting review helpfulness and their classification. Namvar [6] used Neural Network (NN) along with cross-validation techniques to classify new reviews as either helpful or unhelpful. Zhang and Lin [19] used Support Vector Machines (SVM) to classify review helpfulness and adopted the linear regression technique in helpfulness prediction for non-English reviews. Singh et al. [20] proposed that ensemble learning techniques perform better than linear regression techniques in predicting helpfulness ratio. In the study of Lee et al. [21], they applied Decision Tree, Random Forest, Logistic Regression, and SVM to predict the helpfulness of hotel reviews in the U.S. market. Table 1 summarizes the variables and techniques used in the literature of online reviews.

**Table 1. Factors influencing the helpfulness of online reviews**

| Reference | Study variables   | Data source        | Data size | Classification techniques  |
|-----------|---|--------------------|-----------|--|
| [6]       | Engagement, recency, frequency, order, sentiment, score, length           | Amazon.com         | 4,675     | Neural network   |
| [13]      | TF-IDF, SVD, rating, date, writing style, semantic, length                | Download.cn et.com | 3,460     | Ordinal logistic regression model                                |
| [19]      | Engagement, reviewer fan number, length, sentiment, product review number | Yelp.com           | 4,248     | Support vector machines, linear and negative binomial regression |
| [20]      | Polarity, subjectivity, entropy, reading ease, length, rating             | Amazon.com         | 622,494   | Gradient boosting algorithm                                      |
| [21]      | Review quality, sentiment, and reviewer characteristics                   | TripAdvisor.com    | 1,170,246 | Random Forest, logistic regression & SVM                         |
| [22]      | Review extremity, review depth, product type, rating                      | Amazon.com         | 1,587     | Tobit Regression   |
| [23]      | Review rating, order, time, reviewer-level average rating                 | Amazon.com         | 74,657    | Logistic Regression  |
| [24]      | Ratings, sentiments, reliability, writing style                           | Amazon.com         | 610,713   | Natural language processing and regression analysis              |

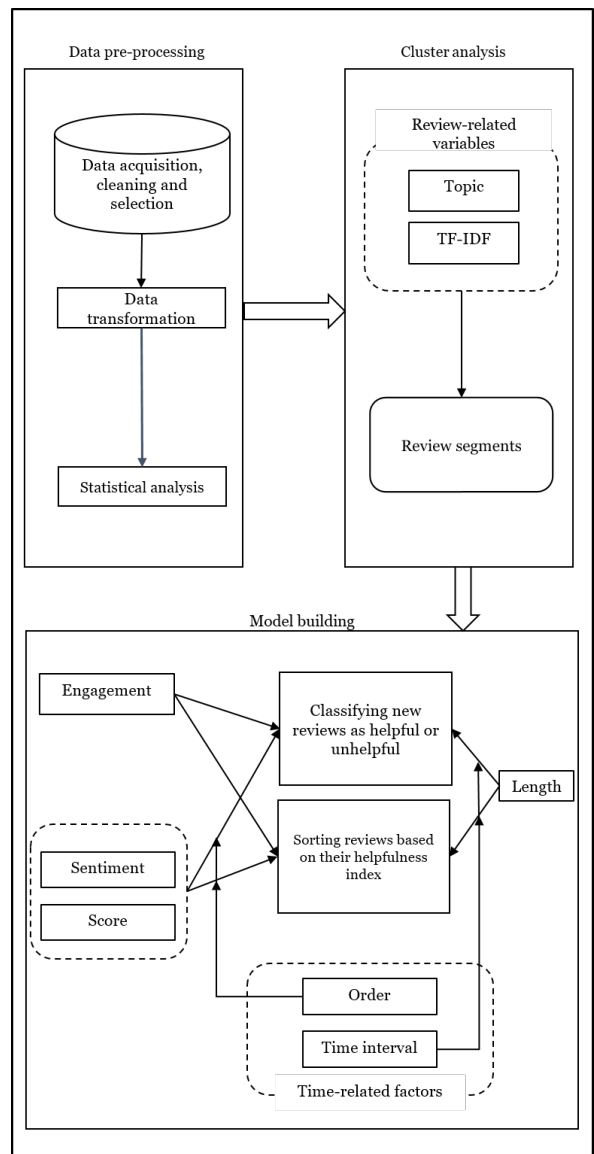
Based on the literature discussed above, this study uses three sets of variables to develop models for predicting the online review helpfulness, namely review-, reviewer-, and time-related variables. We also developed a binominal variable for the review helpfulness (see Table 2). Review-related variables are score, sentiment, length, TF-IDF, and topic. We explain how we measure them in section 4.2. Time-related variables are order and time interval, and there is one reviewer-related variable, engagement. All the variables, except engagement, are at the review level. If a reviewer posts more than one review, his or her engagement measure would be applied to all the reviews he or she has written. This approach to incorporate engagement in studies online reviews is in line with the previous studies (e.g., [5]).

**Table 2. Description of the study variables**

| Study variables    | Definition   |
|--------------------|--|
| Engagement         | The number of past reviews posted by a reviewer [19].  |
| Order              | The total number of reviews written on the product before the given review [15].   |
| Time Interval      | The number of days between the day the review was written on the product, and the day when the previous review was written on the given product [2]. |
| Score              | The numeric (i.e., between 1-5) valence or star rating of a review [17].   |
| Sentiment          | The emotion that customers feel in a review toward the product and can vary from positive to neutral to negative [19].                               |
| Length             | The number of words in a review [17].  |
| TF-IDF             | Term Frequency-Inverse Document Frequency presents the importance of each word across the corpus using numbers [25].                                 |
| Topic              | Topic represents a pattern of keywords embedded in review comments [26].   |
| Review helpfulness | The ratio of helpful votes to total votes on review and indicates whether a review has been helpful or not [8].                                      |

### 3. Research method

We adopt a structured knowledge discovery process [27, p. 120] to classify and predict the helpfulness of online reviews. Figure 1 shows the research process consisting of three main steps: data pre-processing, review clustering, and model building.



**Figure 1. Research method followed**

We use a dataset of online reviews and follow three steps for data pre-processing, including data cleaning and selection, data transformation, and statistical analysis. Applying deep learning techniques [28] for transforming unstructured text data into a structured format, we operationalize the study variables.

In the second step, we use cluster analysis to find review clusters that encompass observations that are like one another and dissimilar to those observations in other clusters [27, p. 443]. To develop clusters that have small inter-point distances in relation to the distance to observations in other clusters, we first identify ideal variables for cluster analysis and then using an appropriate clustering technique, we determine the ideal number of clusters.

In the last step, we train a dataset of online reviews using classification algorithms to label online reviews. For each review, the classifier learns from a pre-processed set of online reviews with the assigned label. We use cross-validation techniques to avoid overfitting and achieve more accurate measures for classification performance [27, p. 370].

In the model building section, we incorporate two time-related variables, namely time interval and order, as moderators. The use of the time interval as a moderator is based on the signaling theory [29], in which provided signals can be impacted by information overload. We argue that when there are fewer reviews on a given product there is a higher time interval for subsequent review as there is a time gap between reviews. In such circumstances, information overload is unlikely to occur. Consequently, readers would be interested in detailed (lengthy) review as they provide more information on the product compared to shorter reviews. As a result, we use the time interval as a moderator for review length.

The use of the order as a moderator is based on the social influence theory [30]. Under the social influence of prior reviewers, subsequent reviewers tend to adjust their product evaluations when writing their own reviews to conform to prior opinions [13, 31-33]. Motivated by the issue, we argue that when review order is less, there is less social influence. Consequently, review score and sentiment are less biased and impacted by the previous reviews.

## 4. Data pre-preparation and measure development

### 4.1. Data acquisition, cleaning, and selection

We extracted a dataset from Amazon.com (<https://s3.amazonaws.com/amazon-reviews-pds/tsv/index.txt>) with more than 130 million reviews from 1995 to 2015 on 46 product categories. We then narrowed down to only five different product categories, which are the most popular products, resulting in 21.177 million reviews, of which 10% were grocery products, 12% were pet-products, 23% were sports products, 24% were beauty products, 29% were home products.

We then only chose reviews from verified purchases resulting in 18,311,121 reviews. Because of methodological requirements and the need for considering order and time interval (see Section 3), we only considered reviews written on the products with more than ten reviews. We further limited our analysis to the products which had at least a review in 2015. Our final dataset consisted of 247,715 reviews on 12,817 products from 229,214 reviewers.

### 4.2. Data transformation

We used the cleaned and reduced dataset to develop 8 study variables (see Table 2). We used 6 variables for predictive model building and two variables for cluster analysis (see Figure 1). We explain the operationalization of these variables next.

Engagement is calculated by the number of reviews posted by a reviewer [41].

Order of a review is the total number of reviews written on the product before the given review. As the number of reviews on different products varies from 10 to 1822, we divided order by the total number of reviews written on the given product to standardize it in the range of 0 to 1. The lowest value for standardized order shows the earliest review, and the highest value, that is 1, shows the most recent review.

Time interval of a review is the number of days between the day the review was written on the product, and the day when the previous review was written on the given product.

Score is the rating provided by a verified purchaser to the product when writing the review.

Review helpfulness (%) is the total number of helpful votes on the given review divided by the total votes. We use this index to develop the predictive model.

Review helpfulness (label): We label a review as helpful if its helpfulness index is more than 60% [42]. We use this label to develop the classification model. The dataset contains more than 71% of unhelpful reviews and 28% of helpful reviews.

Sentiment: to calculate sentiment, we first used the VADER package in Python to obtain an initial positive or negative identifier. Utilizing the deep learning algorithm capability from Keras package, we used the 'Adam' optimizer with 'binary cross-entropy' [34]. We set the model to 40 epochs and 32 batch sizes to recursively run through data. This model was then applied over each review's text to predict the sentiment score.

We used the variables mentioned above in the predictive model building. We develop two other variables for clustering, which we explain next. We applied TF-IDF to our set of tokens to produce a numerical reflection of how important words are and then distributed throughout the corpus. TF-IDF indicates the importance of each word across the corpus using numbers [25]. The weights of words increase analogously with the number of occurrences in the corpus [40]. We then reduced the high dimensionality of TF-IDF outcomes using an SVD (Singular Value Decomposition) algorithm [35] to abstract data to just three dimensions. The resulting variables are names SVD1, SVD2, and SVD3.

We applied topic modeling using deep learning methods. The developed topics signify information about reviews as they are considered as a repeated pattern of keywords in a document [26]. We used the Doc2vec algorithm to provide a vector representation of topics embedded in review comments. We used a Distributed Bag of Words (DBOW) method to convert the cleaned text to a vector matrix representing reviews. The algorithm produced a numerical representation of comments as an array of fixed length vectors regardless of the length of each comment. The comments expressing similar topics are demonstrated with very close vectors. Because of high data dimensionality, we then applied a T-distributed Stochastic Neighbor Embedding (TSNE) to compress all vectors into two dimensions, named Topic1, and Topic 2.

### 4.3. Statistical analysis

We first conducted a correlation analysis (see Table 3). The highest correlation is between sentiment and score at 0.457. This correlation is also observed in other research [30]. As our analysis did not indicate any high correlation, we did not remove any of the study variables.

Table 4 presents the summary statistic of the predictor variables which will be used in the model building step. We did not show the two other sets of variables for clustering as there are reduced variables, and their summary statistics does not provide more information.

**Table 3. Correlation analysis of the predictors**

|    | Attribute     | 1      | 2      | 3      | 4      | 5      | 6      | 7      | 8      | 9     | 10    | 11 |
|----|---------------|--------|--------|--------|--------|--------|--------|--------|--------|-------|-------|----|
| 1  | Engagement    | 1      |        |        |        |        |        |        |        |       |       |    |
| 2  | Order         | -0.019 | 1      |        |        |        |        |        |        |       |       |    |
| 3  | Time Interval | 0.030  | -0.195 | 1      |        |        |        |        |        |       |       |    |
| 4  | Score         | 0.045  | 0.016  | -0.007 | 1      |        |        |        |        |       |       |    |
| 5  | Sentiment     | -0.001 | 0.001  | -0.002 | 0.457  | 1      |        |        |        |       |       |    |
| 6  | Length        | 0.084  | 0.006  | -0.033 | -0.055 | -0.007 | 1      |        |        |       |       |    |
| 7  | SVD1          | 0.001  | -0.002 | 0.003  | 0.001  | 0.241  | -0.001 | 1      |        |       |       |    |
| 8  | SVD2          | 0.002  | 0.000  | 0.001  | -0.003 | 0.046  | -0.002 | -0.427 | 1      |       |       |    |
| 9  | SVD3          | -0.005 | 0.003  | -0.005 | 0.001  | 0.000  | -0.002 | -0.259 | -0.010 | 1     |       |    |
| 10 | Topic1        | -0.009 | 0.003  | -0.028 | -0.010 | -0.002 | 0.017  | 0.002  | 0.002  | 0.000 | 1     |    |
| 11 | Topic2        | 0.004  | -0.020 | -0.022 | -0.008 | 0.000  | 0.002  | 0.001  | -0.002 | 0.000 | 0.014 | 1  |

**Table 4. Descriptive statistics of the structured variables**

|               | Min   | Max  | Mean  | SD    |
|---------------|-------|------|-------|-------|
| Engagement    | 1     | 1319 | 7.83  | 18.37 |
| Order         | 1     | 1822 | 39.45 | 76.09 |
| Time Interval | 0     | 201  | 6.76  | 11.08 |
| Score         | 1     | 5    | 4.19  | 1.30  |
| Sentiment     | -0.99 | 0.99 | 0.75  | 0.55  |
| Length        | 1     | 386  | 35.05 | 53.70 |

## 5. Results

As shown in Figure 1, after data pre-processing, the proposed research undertakes three machine learning approaches 1) review clustering, 2) classifying new reviews, and 3) sorting reviews based on their helpfulness. We explain each step next.

### 5.1 Review clustering

We used topics and SVDs, and applied k-Means algorithm to develop review clusters. This method uses the value of k, as determined by the user, to make a k number of clusters. First, we used an initial value for k by using the square root of the total number of records divided by two [27, p. 451].

Attempting to find the optimal number of clusters, we used the elbow method to adjust the value for k. We evaluated the clusters using the Davies-Bouldin Index [36], which assesses intra-cluster similarity and inter-cluster differences. This method measures the average distance between the center of a cluster and the objects it contains. k-Means desires the lower value of this index as lower values indicates a more precise grouping of records and higher differentiation between individual clusters.

**Table 5. Elbow method in choosing the cluster number**

|                    |      |      |      |      |
|--------------------|------|------|------|------|
| Number of clusters | 2    | 3    | 4    | 5    |
| DB score           | 1.23 | 0.91 | 0.87 | 0.92 |

We chose K=4 for the k-Means algorithm as it shows the lowest DB index. After clustering, using K=4, the analysts resulted in 59231 reviews in cluster 1, 57,936 reviews in cluster 2, 60,481 reviews in cluster 3 and 61,649 reviews in cluster 4.

## 5.2 Classifying new reviews

In training each classifier, we used A) Neural Network (NN), B) Support Vector Machine (SVM), and C) Deep learning (DL). We used the Review helpfulness (label) (see Section 4.2) as a class (target) variable and used moderated length, moderated sentiment, moderated score, and activity as predictors in each cluster.

We implemented all models and calculations in Python. For model A (SVM), we used the C-Support Vector Classification within the Sci-kit Learn package. For model B (NN), we used the MLPClassifier within Sci-Kit Learn package. We set the topology to three hidden layers (12, 8, 1) with 1000 iterations due to its performance compared to other topologies. For model C (DL), we used the Keras deep learning API. We used the sequential class, which groups linear stacks (0 = negative and 1 = positive) into stacks. We incorporated three densely connected Neural Network layers (12,8,1) in these models with the activations (relu, relu, sigmoid), and used 40 epochs (one pass over the entire dataset) to improve and train the model using a batch size of 50. We provided the cross-entropy loss between true labels and predicted labels (binary\_crossentropy) and optimized it by the Adam algorithm.

The performance of this model is tested using a 5-fold cross-validation method. We first compared three approaches in model building using ROC (see Table 6).

**Table 6. Comparing ROC of the proposed classifiers (%)**

|           | Model A<br>SVM | Model B<br>NN | Model C<br>DL |
|-----------|----------------|---------------|---------------|
| Cluster 1 | 53             | 53            | 62            |
| Cluster 2 | 52             | 50            | 59            |
| Cluster 3 | 53             | 54            | 58            |
| Cluster 4 | 53             | 50            | 60            |

We chose Deep Learning (Model C) for the rest of our analysis due to its higher ROC value compared to the other techniques. We computed the classic evaluation metric, accuracy, and compared the results

with three seminal models (See Table 7). These models are 1) Foreman et al.'s model [37] 2) Mudambi and Schuff's model [22] 3) Salehan and Kim's model [38]. The results of our model building and cross-validation indicate that the proposed models outperform the existing models for predicting the helpfulness of online reviews.

**Table 7. Classification accuracy (%)**

|                          |      |
|--------------------------|------|
| Cluster 1                | 72.3 |
| Cluster 2                | 72.2 |
| Cluster 3                | 71.5 |
| Cluster 4                | 71.7 |
| Proposed Model (average) | 71.9 |
| Foreman et al.           | 57   |
| Mudambi & Schuff         | 68   |
| Salehan & Kim            | 61   |

Although our proposed model shows superiority in terms of accuracy compared to the seminal models in the literature, it did not show acceptable values for recall and F-score. To improve all the performance metrics, we balanced our dataset using the under-sampling technique. Table 8 compares the performance of our proposed approach when using a balanced and imbalanced dataset.

**Table 8. Classification accuracy (%)**

| Data       | Accuracy | Recall | F-score |
|------------|----------|--------|---------|
| Imbalanced | 71.9     | 30     | 10      |
| Balanced   | 60       | 76     | 64      |

## 5.2 Sorting reviews based on their helpfulness

We adopted RMSProp algorithm to predict the helpfulness index of the new reviews. The predicted helpfulness can be used to sort the reviews. We applied the conceptual model in Fig 1 to develop the prediction model. The numerical helpfulness metric is chosen (see Section 4.2) as the target variable and we selected moderated length, moderated sentiment, moderated score, and activity as predictors in four developed clusters. Table 9 shows the error rate in predicting the helpfulness index.

**Table 9. Error rate in predicting the helpfulness index**

| RMSE | MSE  | MAE  |
|------|------|------|
| 0.41 | 0.16 | 0.34 |

## 6. Discussion

To predict review helpfulness and to evaluate the practical relevance of our model, it should be assessed whether the proposed model achieves the main goal: identifying helpful online product reviews to display

them first. This assessment should correspond to the main goal of online retailers. In contrast to previous works that studied helpfulness based on review-related factors, this study additionally incorporates the time-related dimension of customer reviews and allows for a deeper understanding of the assessment of consumer reviews on online retailer platforms.

As the main finding, this study provides strong evidence that the review order and time interval play key roles in predicting the online review helpfulness. It is in line with findings of previous research that later reviews can be more helpful than what the retailer or the voting system shows, and recent reviews can be less helpful than what the system shows. One possible explanation can be that for products with several reviews, readers only refer to either the reviews with the most (helpful) votes or recent reviews. However, reviews which are not recent or old will remain unread and will not receive any helpfulness vote.

This study not only considers the impacts of time-related factors on the helpfulness of online reviews but also uses review- and reviewer-related factors on the helpfulness of online customer reviews. In line with the previous works on online reviews, this study confirms the impact of sentiment and length of reviews, the score provided by the reviewer on a product, and the activity of reviewer on the helpfulness of online reviews.

The order of reviews moderates the impacts of sentiment and score of reviews and their helpfulness. One explanation could be that reviews posted late are more likely to have lower score, whereas reviews made early usually present higher satisfaction. Skeptical customers would read more reviews. They would very likely sort reviews by recency and read the ones are shown first. The other explanation, yet to be tested, is that if a new product is just introduced in the market by an online retailer, high score review spamming may occur on the product.

Time interval of reviews moderates the impact of review length on the helpfulness of online reviews. The analysis shows if reviews are posted within a few days after the previous review on a given product (lower time interval for the review), it is more likely that readers are already overloaded by too much review on a given product. In such cases, a product attracts several reviews and the readers of reviews on the given product might be overloaded by too much information. Whereas when reviews have high time intervals, it means there has been a time gap between the previous review and the new review on a given product. In such cases, it is more likely that readers are not overloaded by too much information and as a result they are interested in more elaboration on the details, and lengthy reviews are more desirable.

Related to the practical relevance shown by the predictive evaluation (see Tables 7), the proposed model in this study significantly outperforms the accuracy of benchmarking models when classifying online reviews as helpful or unhelpful. The 71.9% accuracy of the proposed model is, however, questionable as the other performance metrics, e.g., recall, were lower. That said, the proposed model in Table 7 only performed well in predicting unhelpful reviews but does not work well in predicting helpful reviews. An immediate remedy for this issue is using a balanced dataset with an equal number of reviews of each class. The performance of such a model is shown in Table 8. Even though the accuracy is lower than 7, this model performs relatively well in predicting helpful and unhelpful reviews.

Currently, online retailers such as Amazon.com sort reviews based on their recency or helpfulness (voted by readers). The analysis of this study shows that, not all the reviews receive votes from the readers, and the votes received by readers are not comprehensive metrics in sorting online reviews as temporal factors are influential in the received votes. Employing the proposed approach in this study, online retailers can, first, label reviews that have not received any votes with the performance shown in Table 9. Secondly, they can provide a third option for sorting online reviews, in addition to the existing ones namely recency and helpfulness, which is an intelligent helpfulness rank of the online reviews provided by the online review platform.

## 7. Conclusion and future work

Online review platforms provide open, convenient communication channels for sharing and gathering consumer reviews. These online reviews represent unique and valuable information sources for e-commerce firms to understand the perceptions of their customers about their shopping experience. This study proposed a novel approach for predicting the helpfulness of online reviews, which incorporates the impacts of time-related variables in predicting the helpfulness of online reviews from the new reviewers and existing reviewers.

The research reported in this paper sheds light on the understanding of online review helpfulness and the design of a better helpfulness voting mechanism for online review platforms. It advances the understanding of the various variables that have an impact on the helpfulness of online reviews. It also has implications for consumers to leverage online product reviews to infer actual product quality.

The analysis of insights provides benefits to online retailers planning to implement online reviews to

improve their customer experience. Since the helpfulness of reviews is directly related to the purchase intention of potential customers, the findings can help companies to enhance their communication strategies regarding product descriptions, social media content and advertisement. Currently, most of the online retailers, such as Amazon.com, order online customer reviews based on the number of helpfulness votes or their recency. One possible contribution of this study is using the proposed method when not enough customers vote for the reviews of a newly launched product. In this context, it should not be assumed that positive or negative reviews are generally perceived as more helpful. Instead, the role of review ratings in relation to perceived helpfulness also depends on the time the review is posted.

By incorporating the moderating impacts of time-related variables and applying deep learning techniques, this study outperforms the existing benchmarking models for predicting the helpful reviews. Future studies still need to be conducted to identify when and how fake reviews are written. Also, this study examined the proposed approach using a dataset of online review about products from an online retailer; future works can examine this approach in other e-commerce platforms which collect reviews regarding services rather than products, as the discrepancies in the helpfulness of service reviews are more than a product review. Finally, even though we utilized a solid support from literature to use 60% as the appropriate cutting point to label helpful reviews, in the future works we can examine other cutting points (e.g., [39]) for sensitivity analysis in robustness check of our findings.

## 8. References

- [1] S. Park and J. L. Nicolau, "Asymmetric effects of online consumer reviews," *Annals of Tourism Research*, vol. 50, pp. 67–83, 2015.
- [2] H. Hong, D. Xu, G. A. Wang, and W. Fan, "Understanding the determinants of online review helpfulness: A meta-analytic investigation," *Decision Support Systems*, vol. 102, pp. 1–11, 2017.
- [3] "Local Consumer Review Survey: How Customers Use Online Reviews", BrightLocal, 2019. [Online]. Available: <https://www.brightlocal.com/research/local-consumer-review-survey/>. [Accessed: 13- Jul- 2020].
- [4] S. Banerjee, S. Bhattacharyya, and I. Bose, "Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business," *Decision Support Systems*, vol. 96, pp. 17–26, Apr. 2017
- [5] M. Siering, J. Muntermann, and B. Rajagopalan, "Explaining and predicting online review helpfulness: The role of content and reviewer-related signals," *Decision Support Systems*, 2018.
- [6] M. Namvar, "A Novel Approach to Predict the Helpfulness of Online Reviews," presented at the Hawaii International Conference on System Sciences, 2020.
- [7] X. Sun, M. Han, and J. Feng, "Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products," *Decision Support Systems*, vol. 124, p. 113099, Sep. 2019.
- [8] A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, Oct. 2011.
- [9] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3751–3759, May 2015.
- [10] M. S. I. Malik and A. Hussain, "Helpfulness of product reviews as a function of discrete positive and negative emotions," *Computers in Human Behavior*, vol. 73, pp. 290–302, Aug. 2017.
- [11] A. Y. K. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Computers in Human Behavior*, vol. 54, pp. 547–554, Jan. 2016.
- [12] T. L. Ngo-Ye and A. P. Sinha, "The influence of reviewer engagement characteristics on online review helpfulness: A text regression model," *Decision Support Systems*, vol. 61, pp. 47–58, 2014.
- [13] Q. Cao, W. Duan, and Q. Gan, "Exploring determinants of voting for the 'helpfulness' of online user reviews: A text mining approach," *Decision Support Systems*, vol. 50, no. 2, pp. 511–521, 2011.
- [14] B. Lutz and N. PrÄ, "The Longer the Better? The Interplay Between Review Length and Line of Argumentation in Online Consumer Reviews," p. 9.
- [15] J. Qi, Z. Zhang, S. Jeon, and Y. Zhou, "Mining customer requirements from online reviews: A product improvement perspective," *Information & Management*, vol. 53, no. 8, pp. 951–963, Dec. 2016.
- [16] R. Thakur, "Customer engagement and online reviews," *Journal of Retailing and Consumer Services*, vol. 41, pp. 48–59, Mar. 2018.
- [17] A. Y. K. Chua and S. Banerjee, "Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth: Understanding Review Helpfulness as a Function of Reviewer Reputation, Review Rating, and Review Depth," *J Assn Inf Sci Tec*, vol. 66, no. 2, pp. 354–362, Feb. 2015.
- [18] S. Lee and J. Y. Choeh, "The determinants of helpfulness of online reviews," *Behaviour & Information Technology*, vol. 35, no. 10, pp. 853–863, Oct. 2016.
- [19] Y. Zhang and Z. Lin, "Predicting the helpfulness of online product reviews: A multilingual approach," *Electronic Commerce Research and Applications*, vol. 27, pp. 1–10, Jan. 2018.
- [20] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. Kumar Roy, "Predicting the 'helpfulness' of online consumer reviews," *Journal of Business Research*, vol. 70, pp. 346–355, Jan. 2017.
- [21] P.-J. Lee, Y.-H. Hu, and K.-T. Lu, "Assessing the helpfulness of online hotel reviews: A classification-based



approach,” *Telematics and Informatics*, vol. 35, no. 2, pp. 436–445, May 2018.

[22] S. M. Mudambi and D. Schuff, “Research note: What makes a helpful online review? A study of customer reviews on Amazon.com,” *MIS Q.*, pp. 185–200, 2010.

[23] D. Godes and J. Silva, “Sequential and Temporal Dynamics of Online Opinion”, *Marketing Science*, vol. 31, no. 3, pp. 448-473, 2012.

[24] N. Hu, I. Bose, N. Koh and L. Liu, “Manipulation of online reviews: An analysis of ratings, readability, and sentiments”, *Decision Support Systems*, vol. 52, no. 3, pp. 674-684, 2012.

[25] Y. Zhou, S. Yang, Y. Li, Y. Chen, J. Yao and A. Qazi, “Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining”, *Information Processing & Management*, vol. 57, no. 2, p. 102179, 2020.

[26] M. Yang, Q. Qu, Y. Shen, K. Lei and J. Zhu, “Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning”, *Neural Computing and Applications*, vol. 32, no. 11, pp. 6421-6433, 2018.

[27] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*, 3rd ed. Elsevier, 2011

[28] H. Yi, H. Jung, and S. Bae, “Deep Neural Networks for traffic flow prediction,” 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), 2017.

[29] M. Siering, J. Muntermann and B. Rajagopalan, “Explaining and predicting online review helpfulness: The role of content and reviewer-related signals”, *Decision Support Systems*, vol. 108, pp. 1-12, 2018.

[30] S. Zhou and B. Guo, “The Interactive Effect of Review Rating and Text Sentiment on Review Helpfulness,” *Lecture Notes in Business Information Processing E-Commerce and Web Technologies*, pp. 100–111, 2015.

[31] Y. Chen and J. Xie, “Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix”, *Management Science*, vol. 54, no. 3, pp. 477-491, 2008.

[32] Z. Chen and N. Lurie, “Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth”, *Journal of Marketing Research*, vol. 50, no. 4, pp. 463-476, 2013.

[33] C. Cheung and M. Lee, “What drives consumers to spread electronic word of mouth in online consumer-opinion platforms”, *Decision Support Systems*, vol. 53, no. 1, pp. 218-225, 2012.

[34] A. Nagisetty and G. Gupta, “Framework for Detection of Malicious Activities in IoT Networks using Keras Deep Learning Library - IEEE Conference Publication”, [Ieeexplore.ieee.org](http://Ieeexplore.ieee.org), 2020.

[35] M. Moonen, P. V. Dooren, and J. Vandewalle, “A Singular Value Decomposition Updating Algorithm for Subspace Tracking,” *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 4, pp. 1015–1038, 1992.

[36] U. Maulik and S. Bandyopadhyay, “Performance evaluation of some clustering algorithms and validity indices,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, no. 12, pp. 1650–1654, 2002.

[37] B. Dervin and L. Foreman-Wernet, “Sense-Making Methodology as an Approach to Understanding and Designing for Campaign Audiences,” in *Public communication campaigns*, R. E. Rice and C. K. Atkin, Eds. Sage, 2012, pp. 147–162.

[38] M. Salehan and D. Kim, “Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics”, *Decision Support Systems*, vol. 81, pp. 30-40, 2016.

[39] X. Sun, M. Han, and J. Feng, “Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products,” *Decision Support Systems*, vol. 124, p. 113099, Sep. 2019.

[40] B. Bansal and S. Srivastava, “Hybrid attribute based sentiment classification of online reviews for consumer intelligence”, *Applied Intelligence*, vol. 49, no. 1, pp. 137-149, 2018.

[41] R. Barbado, O. Araque and C. Iglesias, “A framework for fake review detection in online consumer electronics retailers”, *Information Processing & Management*, vol. 56, no. 4, pp. 1234-1244, 2019.

[42] R. Cangelosi and A. Goriely, “Component retention in principal component analysis with application to cDNA microarray data,” *Biol. Direct*, vol. 2, no. 1, p.2, 2007.

[43] L. Li, T. Goh and D. Jin, “How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis”, *Neural Computing and Applications*, vol. 32, no. 9, pp. 4387-4415, 2018.

[44] M. Luca and G. Zervas, “Fake it Till You Make it: Reputation, Competition, and Yelp Review Fraud”, *SSRN Electronic Journal*, 2013.

[45] P. Wu, “In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews”, *Psychology & Marketing*, vol. 30, no. 11, pp. 971-984, 2013.