

Article

Profiling and Predicting the Cumulative Helpfulness (Quality) of Crowd-Sourced Reviews

Muhammad Bilal ^{1,2,*}, Mohsen Marjani ^{1,2,*}, Ibrahim Abaker Targio Hashem ^{1,2,*}, Abdullah Gani ^{3,4}, Misbah Liaqat ⁵ and Kwangman Ko ^{6,*}

- ¹ School of Computing and IT, Taylor's University, Subang Jaya 47500, Malaysia
- ² Centre for Data Science and Analytics (C4DSA), Taylor's University, Subang Jaya 47500, Malaysia
- ³ Department of Computer System and Technology, University of Malaya, Kuala Lumpur 50603, Malaysia; abdullah@um.edu.my
- ⁴ Faculty of Computing and Informatics, University Malaysia Sabah, Labuan International Campus, Labuan 87000, Malaysia
- ⁵ Department of Computer Science, Air University, Islamabad 44000, Pakistan; misbahliaqat@gmail.com
- ⁶ Department of Computer Engineering, Sangji University, Wonju 220-702, Korea
- Correspondence: mbilal.csit@gmail.com (M.B.); marjanimohsen@gmail.com (M.M.); IbrahimAbaker.TargioHashem@taylors.edu.my (I.A.T.H.); kkman@sangji.ac.kr (K.K.)

Received: 27 August 2019; Accepted: 3 September 2019; Published: 24 September 2019



Abstract: With easy access to the Internet and the popularity of online review platforms, the volume of crowd-sourced reviews is continuously rising. Many studies have acknowledged the importance of reviews in making purchase decisions. The consumer's feedback plays a vital role in the success or failure of a business. The number of studies on predicting helpfulness and ranking reviews is increasing due to the increasing importance of reviews. However, previous studies have mainly focused on predicting helpfulness of "reviews" and "reviewer". This study aimed to profile cumulative helpfulness received by a business and then use it for business ranking. The reliability of proposed cumulative helpfulness for ranking was illustrated using a dataset of 1,92,606 businesses from Yelp.com. Seven business and four reviewer features were identified to predict cumulative helpfulness using Linear Regression (LNR), Gradient Boosting (GB), and Neural Network (NNet). The dataset was subdivided into 12 datasets based on business categories to predict the cumulative helpfulness. The results reported that business features, including star rating, review count and days since the last review are the most important features among all business categories. Moreover, using reviewer features along with business features improves the prediction performance for seven datasets. Lastly, the implications of this study are discussed for researchers, review platforms and businesses.

Keywords: review platforms; crowd-sourced reviews; profiling helpfulness; ranking businesses; helpfulness prediction

1. Introduction

The rapid growth of the Internet and the popularity of crowd-sourced review platforms have introduced electronic Word-of-Mouth (e-WoM) communities that provide a massive amount of User-Generated Content (UGC), i.e., online product reviews [1,2]. The popular review websites, e.g., Yelp, Amazon, TripAdvisor, IMDB, Yahoo, Google, etc., serve as an essential source of information and help users in evaluating product quality and making purchase decisions [3–6]. These websites, despite differing, i.e., Yelp reviews business, Amazon is an e-commerce website and review products, TripAdvisor is a booking website, etc., the principle of review helpfulness are common [7]. According to

Bright Local [8], 86% of consumers read online reviews, whereas 91% of consumers trust online reviews. The volume of online review is increasing day by day. Currently, there are more than 730 million reviews on TripAdvisor [9] and more than 184 million on Yelp [10].

The colossal quantity of unstructured data generated by e-WOM communities has become a source of big data to study real consumer behavior [11–13], which also introduced many challenges for both businesses and consumers [14]. The "review helpfulness" is an important dimension of online reviews, which shows the subjectivity and quality perceived by the crowd [15,16]. To overcome the problem of information overload and facilitate the consumers in finding helpful reviews from thousands of confusing reviews several solutions have been proposed using statistical modelling and Machine Learning (ML) [17–19].

The topic of predicting helpfulness of reviews has been studied by many researchers using similar features but reported inconsistent and contradictory results regarding the performance of different features in predicting helpfulness [20]. Most of the solutions introduced by previous studies were for a specific category, product or platform [21,22]. Researchers have tried to propose a generalized solution for different review platforms and product categories by using only textual features in making the prediction. However, they also suggested utilizing reviewer and product features to enhance the prediction performance [23]. The datasets used for predicting helpfulness by previous studies are mostly different, small size and overdispersed [21,24]. Diaz and Ng [7] highlighted the disorganized status of research in the area of review helpfulness prediction.

The quality of a business is represented by the average star rating of all reviews. Similarly, the quality of reviews received by a product is reflected by the helpfulness of reviews based on their information value. By using the review helpfulness as a tool, customers' ability to access the quality of a product or business has been greatly improved. The helpful user reviews are of great use to potential consumers as they provide information about the quality of a product that helps in evaluating and making purchase decisions [25–27]. The businesses with more useful reviews are likely to attract more customers and the increased revenue in comparison to businesses with less useful reviews [28–30].

The review website usually allows the reader to give feedback to a review, i.e., helpful/not-helpful vote on Amazon, Useful vote on Yelp. This simple feedback boosted Amazon revenue by \$2.3 billion [31]. Lu et al. [32] reported that a major portion of the reviews has very few or no useful votes because the latest reviews did not get enough time to receive useful votes. Hence, the useful votes for individual reviews are too sparse to access the quality of reviews received by a product [33,34]. There is a huge volume of reviews even for a single business, and it is challenging to see the quality reviews received by the business. Moreover, the quality of reviews received by one business is different from others, even for the same category. Therefore, similar to the average star rating of the business, the cumulative helpfulness of reviews for a business should be calculated as well. The cumulative helpfulness can be calculated from the perspective of the reviewer as well as business. "Cumulative helpfulness" is the total helpful votes received by all reviews for a specific business or written by a particular reviewer.

Due to the importance of review helpfulness, the number of studies trying to explore the helpfulness of crowd-sourced reviews is continuously increasing. Despite these rising numbers, the majority of studies have explored the helpfulness of reviews for limited categories, e.g., shopping, restaurants, etc., and platforms, e.g., Amazon.com, while ignoring reviews categories, i.e., travel, hotel, health, and platforms, i.e., Yelp and TripAdvisor [35]. In addition, the researchers have proposed many statistical and ML models for (i) predicting helpfulness of "review" and "reviewer"; and (ii) finding and ranking the top-k helpful "review" and "reviewer". However, according to our knowledge, there is no published research article attempting to find and predict the cumulative helpfulness (quality) of reviews for a business.

Therefore, the cumulative helpfulness of reviews received by a "business" still needs to be investigated. To fill the gap, this study aimed at finding the cumulative helpfulness of reviews received by a business and compared the prediction performance of various ML algorithms on datasets of

different size and business categories. The main contributions of this paper are summarized as follows: (a) propose and calculate the cumulative helpfulness of reviews received by a business; (b) rank and compare top k businesses using cumulative helpfulness, review count and star rating; (c) identify and operationalize the business and reviewer features for predicting cumulative helpfulness of reviews received by a business; (d) analyze the performance of various learning algorithms to predict the cumulative helpfulness of reviews for a business using datasets of different size and business categories; (e) examine the impact of reviewer features in predicting cumulative helpfulness of a business; and (f) explore the importance of different business and reviewer features for predicting cumulative helpfulness.

The rest of the paper is organized as follows. Section 2 gives a brief overview of literature related to online review helpfulness prediction. Section 3 illustrates the research methodology. Section 4 reports and discusses the experimental results. Section 5 discusses the implications. Section 6 outlines the limitations and future work. Finally, Section 7 concludes the study.

2. Literature Review

The literature on predicting helpfulness of reviews is continuously increasing as it becomes a critical factor for consumers in making purchase decisions [20,36]. This section provides an overview of the current state of the literature on predicting helpfulness and ranking reviews using multiple features, i.e., review content, reviewer, product/business, emotions, etc., and various techniques. A study found that review extremity, depth, and type of product affect the perceived helpfulness of reviews by analyzing data collected from Amazon.com. The type of product plays the role of moderator between depth and helpfulness [25]. Cao et al. [15] studied the relation of review features with helpfulness. It was found that reviews with extreme opinions are more helpful when compared with neutral reviews. A study explored the helpfulness of online reviews by using both qualitative, i.e., reviewer experience, and quantitative, i.e., word count, features. The analysis was performed on 1375 reviews and data of the top-ranking 60 reviewers from Amazon.com. The relation of the length of review with the helpfulness appeared significant up to a certain threshold. In addition, the reviewer experience had reported no significant relation with helpfulness. However, the past record of reviewer helpfulness can predict future helpfulness. The study reported a changing impact of different review and reviewer related features on perceived helpfulness [37]. The important reviewer features, along with review features, were examined. Performance of popular ML algorithms. i.e., NNet, Random Forest (RandF), Stochastic GB, etc., were compared by performing analysis over three datasets containing 32,434, 109,357, and 59,188 reviews collected from Amazon.com. The proposed review content-related features give the best performance in comparison with reviewer features and previously proposed models. The linguistic features of reviews along with the reviewer helpfulness per day are also strong predictors of review helpfulness [36].

The features that influence review helpfulness prediction were analyzed using review collected from Amazon.com and ML algorithms including Logistic Regression (LGR), Support Vector Regression (SVR), Model tree (M5P) and RandF. The results reported that the relation of different features with the review helpfulness prediction varies for all five categories tested. Moreover, SVR shows the best performance in predicting the review helpfulness for all five categories in comparison with LGR, M5, and RandF. To identify the most helpful review from the massive volume of reviews for a given product or business, a NNet based prediction model was proposed. The results reported the significance of features for predicting helpfulness of reviews [38]. Wu [39], inspired by communication theories, tried to explore the effectiveness of reviews by keeping in consideration review showed the importance of review popularity and helpfulness in evaluating the effectiveness of reviews, reviewer, and product-related features were analyzed using ML algorithms. The data collected from Amazon.com contain 32,434 reviews and 3100 products were analyzed. The results revealed that the proposed review category and reviewer features are better predictors of review helpfulness.

The recency of reviewer, along with the length of activity, also showed statistically significant relation with the helpfulness of reviews [40].

The impact of emotions on the helpfulness of online reviews collect from Amazon.com was studied using Deep Neural Network (DNN). NRC emotion Lexicon was used to extract the emotions attached to reviews. The features that were previously studied, i.e., reviewer, product and linguistics, were used for predicting helpfulness. It was evident from the results that emotions were the best predictors of review helpfulness when features were taken individually. Moreover, the mixture of other features and emotion was reported to produce better overall performance [41]. The relation of review title features with the review helpfulness has been explored by using data for 475 book reviews from Amazon.com. A model was proposed based on review content, reviewer, readability and title features. The proposed model was tested on a collected dataset of book reviews using ML algorithms i.e., Decision Tree (DT) and RandF. It was reported that the review title features were not a significant predictor of review helpfulness [42]. A model based on GB algorithm was proposed to predict review helpfulness by using textual features of reviews, i.e., readability, polarity, and subjectivity. The analysis was performed on reviews related to books, baby products, and electronic products collected from Amazon.in. The results reported that textual features are a better predictor of review helpfulness [19].

Gao et al. [43] studied the consistency and predictability of rating behavior of reviewers over time along with their review helpfulness. The data collected from TripAdvisor.com was analyzed using econometric models. The results reported that the rating behavior of reviewers is consistent over time. Moreover, the reviewers that currently have higher ratings were reported to be more helpful in future reviews. The results were robust when tested over different product categories. The review content and rating were not significantly related, as reported by previous studies. A review helpfulness prediction model was developed by considering the unexplored features. The analysis was performed by collecting 1500 hotel reviews from TripAdvisor.com. The results reported that many notions in review and review type have varying impact on the helpfulness of hotel reviews [44]. The classification of reviews into helpful and not-helpful was performed using 1,170,246 reviews collect from TripAdvisor.com. The ML classification algorithms used include DT, RandF, LGR, and Support Vector Machine (SVM). Accuracy, sensitivity, specificity, precision, recall, and F-measure were used to evaluate the performance. The results reported that the reviewer features were a good predictor for predicting review helpfulness in comparison with review quality and sentiment [45].

Customer reviews from Amazon.in and Snapdeal.com were analyzed using two-layered Convolutional Neural Network (CNN) to predict the most helpful review for a given product. Three filters, namely tri-gram, four-gram, and five-gram, were used to extract the textual features for predicting helpfulness of reviews. As the study relied only on textual features, the proposed approach was reported to be flexible for predicting helpfulness of reviews for any domain. The results showed better performance for the CNN model in comparison with other ML models [23]. The unexplored assumptions, i.e., star rating, equal review visibility, the constant status of review and reviewer, made in previous studies were investigated using data collected from TripAdvisor.com. The review visibility features, e.g., days since the review was posted, days review was displayed on the home page, etc. showed a strong relation with review helpfulness. The M5P showed better performance in comparison with LNR and SVR [35]. Saumya et al. [22] proposed a review ranking approach based on their predicted helpfulness. The features related to review content, reviewer and product were extracted from Amazon.in and Snapdeal.com reviews. RandF was used for the classification of reviews as high-quality and low-quality reviews. Afterwards, GB regressor was used to calculate the helpfulness score of the high-quality review. The top-k reviews were ranked according to the helpfulness score, whereas the low-quality were simply added at the end. The results reported a fair ranking of reviews as the top ten review include few latest reviews along with a few previous reviews.

The impact of review numerical and textual features in predicting review helpfulness were explored by using Amazon reviews. The analysis was performed on the collected data using RandF. It was reported that the numerical features are a significant predictor of review helpfulness for all

three types of reviews, i.e., regular, suggestive and comparative reviews. The review length and complexity were also a significant predictor of helpfulness. However, the relation of review complexity with helpfulness was inverted U-shaped [46]. The effect of user-controlled features, along with other predictors, was investigated using reviews collected from TripAdvisor.com. The results showed varying relation of user-controlled filters with selected features. The Recency, Frequency, and Monetary (RFM) model showed consistency among all controlled variables. Moreover, the rating of review and length were reported as the most important predictors of review helpfulness [47]. The impact of including RFM characteristics of reviewers on the performance of predicting review helpfulness were analyzed using the Bag-of-Words (BoW) model and RFM features produced best results [48]. Mohammadiani et al. [49] divided reviewers into two groups based on their strength of the relationship. The analysis performed on data collected from Epinions.com showed that the effect of review helpfulness on the influence of the reviewer is significant for high similarity.

A study introduced a Deep Learning (DL) model to understand the quality of online hotel reviews. The data collected from Yelp.com and TripAdvisor.com were analyzed using CNN and Natural Language Processing (NLP) to explore the relation of photo provided by the user and review helpfulness. The DL models outperform the other models in predicting helpfulness of reviews. The results reported that the photos provided by the user alone are not a good predictor of review helpfulness. Moreover, combining the photos with the features of review text yielded better performance [50]. The influence of reviewer profile photo on perceived review helpfulness was explored by extracting decorative and information features from photos of 2178 mobile gaming reviews collected from the Google Play store. The experimental results performed using Tobit regression model, reported that the profile photo plays a significant role in the perception of review helpfulness. However, the type of photo did not show any significant impact on review helpfulness. More interestingly, the review length moderates the relation between profile image and review helpfulness rather than review valance or equivocality [51]. The textual features of the review were examined using ML models, i.e., RandF, Naïve Bayes (NB), etc., to identify the quality of hotel reviews available on TripAdvisor. The stylistic features were reported as a more important determinant of review helpfulness, however, by combining stylistic features with content features, produced better prediction results [52].

The language used by reviewers in writing product reviews varies a lot. Four stylistic features were identified and analyzed for their relationship with review helpfulness using data collected from Epinions.com. The stylistic features were reported a good predictor of review helpfulness in comparison with other features. However, it was suggested to use the stylistic features along with social features to gain better performance [53]. Krishnamoorthy [5] proposed a predictive model to investigate the review features that have an impact on reviews helpfulness. The data collected from Amazon.com was analyzed using ML algorithms, i.e., NB, SVM, and RandF. The linguistic features extracted from the review content were analyzed along with readability, subjectivity, and metadata. It was concluded from results that the hybrid set of features produce better accuracy. Moreover, linguistic features were reported as a good predictor for some categories, e.g., books and games. A multilingual technique was introduced to overcome the gap of predicting the review helpfulness for reviews in languages other than English. The dataset of 4248 non-English reviews was collected from Yelp.com. The previously identified features related to review content, business and reviewer were analyzed using regression, i.e., LNR, and classification techniques, i.e., SVM [21]. The analysis of scripts for predicting review helpfulness was performed with the help of human annotators that highlight the important phrases that make a review helpful. The results showed that the script enriched model gives better performance even with small training set in comparison with traditional models, e.g., BoW [54].

This research hypothesized that the cumulative helpfulness prediction using business features, along with reviewer features, give more accurate results and enhance prediction performance. Moreover, the cumulative helpfulness of a business calculated from online reviews can be used

as an alternative to rank businesses efficiently and effectively. This research also explored which ML algorithm gives the best performance and which features are more important in predicting cumulative helpfulness.

3. Research Methodology

In this section, the stages of data collection, problem definition, feature generation and selection, modeling, and evaluation are described in detail. The research methodology of this study is illustrated in Figure 1.



Figure 1. Flow chart of proposed research methodology.

3.1. Data Collection and Pre-Processing

The dataset used in this study was provided by Yelp that spans from 12 October 2004 to 14 November 2018 [55]. The dataset includes information about 192,609 businesses, 6,685,900 reviews, 1,223,094 tips, 200,000 photos, check-in information of 161,950 business. In addition, the dataset also contains information about 1,673,138 users who reviewed the selected business. The dataset contains information of business for 10 metropolitan areas across two countries. The database schema created for the shared dataset is illustrated in Figure 2. This study used information from all sources, excluding tips. To generate a dataset for experimentation, firstly the user's information was mapped across each review. Then, reviews information was grouped and mapped for each business. Along with this, we generated check-in and photo count features for each business. There was a difference of three business after mapping when compared with the actual business count because no reviews were found in the reviews table for those businesses. The label H_m (the cumulative helpfulness of business *m*) was also generated before generating the final dataset of 192,606 businesses. In the final step, we mapped features from review table and business table to final dataset having 192,606 records. The procedure of creating the dataset used in this study is illustrated in Figure 3. Each feature is discussed and described in detail in Section 3.2. The features were normalized using Z-Transformation before being used for predictive modeling. The category of each business was labeled to study its impact along with different features in predicting cumulative helpfulness and analyze the performance of ML models on different sized datasets. We created 11 sub-datasets based on the business category. The information about 12 datasets created and used in this study along with the distribution of businesses by each category is given in Table 1.

Datasets	Size	Distribution (%)
All Categories	192,606	100%
Restaurants	59,318	31%
Shopping	31,235	16%
Home and Local Services	23,098	12%
Other	18,678	10%
Beauty and Fitness	16,525	9%
Health	11,953	6%
Auto	11,757	6%
Arts, Entertainment and Events	9131	5%
Travel and Hotel	4404	2%
Pets	3580	2%
Nightlife	2927	1%

Table 1. Description of datasets and distribution of businesses by category.



Figure 2. Entity relationship diagram of Yelp database.



Figure 3. Flow chart of step performed for creating the dataset.

3.2. Problem Formulation and Model Features

The symbols and variables used in this paper are described in Table 2. This study aimed to profile the cumulative helpfulness of a business. Moreover, this study compared the top-k businesses based on star rating, review count and cumulative helpfulness.

In profiling the cumulative helpfulness for each business, there is a set of businesses $B = \{b_1, b_2, ..., b_m\}$, a set of users $U = \{u_1, u_2, ..., u_n\}$ who write the reviews, and a set of reviews $R = \{r_1, r_2, ..., r_i\}$. H_m denotes the cumulative helpfulness of a business *m* and calculated as in Equation (1), whereas the predicted cumulative helpfulness for business *m* is denoted by \hat{H}_m . $H_{m,i}$ represents the helpfulness of review *i* for business *m*.

B_Stars show star rating, and it ranges from 1 to 5. The average stars received by a business is represented by B_Stars_m and the star rating against a single review is given by B_Stars_{m,i}. B_Stars_{n,i} is the star rating given by a user to review *i. m, n,* and *i* represent the number of businesses, number of users and number of reviews, respectively. The total number of check-ins for a business B_m is denoted by B_Checkin_Count_m. B_Photo_Count_m denotes a total number of photos uploaded for a business B_m. The total number of reviews received by a business B_m is represented by B_Review_Count_m. Moreover, B_Activity_Len_m as in Equation (4), B_First_Review_m as in Equation (2), and B_Last_Review_m as in Equation (3) denote the duration in days between first and last review posted for a business B_m , the duration in days since the first review was posted until the data collection date, and the duration in days since the last was posted until the data collection date, respectively.

The average review count of users for business B_m is represented by U_Review_Count_m as in Equation (5). U_Fans_Count_m as in Equation (6) denotes an average number of fans count for the users who have reviewed business B_m . The average number of user's friends who reviewed business B_m is denoted by U_Friends_Count_m as in Equation (7), whereas U_Compliment_Count_m as in Equation (8) is used for the average number of compliments received by users who reviewed business B_m .

Symbol/Variable	Description
В	a set of businesses
U	a set of reviewers/users
R	a set if reviews
т	business number
п	user number
i	review number
B_m	a business <i>m</i>
U_n	a user <i>n</i>
R_i	a review <i>i</i>
Н	helpfulness
H_m	the cumulative helpfulness of business <i>m</i>
$\hat{\mathbf{H}}_m$	the predicted cumulative helpfulness of business <i>m</i>
H_n	user <i>n</i> cumulative helpfulness
$H_{m,i}$	helpfulness of business <i>m</i> to review <i>i</i>
$H_{n,i}$	helpfulness of user <i>n</i> to review <i>i</i>
B_Stars	star rating
B_Stars_m	average star rating for business m
B_Stars _i	star rating for review <i>i</i>
$B_Stars_{m,i}$	star rating of business <i>m</i> to review <i>i</i>
$B_{\text{Stars}_{n,i}}$	star rating of User <i>n</i> to review <i>i</i>
U_Review_Count _{m,n}	review count of user <i>u</i> who review business <i>m</i>
U_Fans_Count _{<i>m</i>,<i>n</i>}	fans count of user u who review business m
U_Friends_Count _{m,n}	friends count of user u who review business m
U_Compliment_Count _{<i>m</i>,<i>n</i>}	compliment count of user <i>u</i> who reviews business <i>m</i>
B_Review_Count _m	business <i>m</i> review count
B_Checkin_Count _m	business <i>m</i> check-in count
B_Photo_Count _m	business <i>m</i> photo count
B_Activity_Len _m	business m activity length in days
B_First_Review _m	days since first review for business <i>m</i>
B_Last_Review $_m$	days since the last review for business m
U_Review_Count _m	average review count of users for business m
U_Fans_Count _m	average count of user fans for business m
U_Friends_Count _m	average count user friends for business m
U_Compliment_Count _m	average count compliments of users for business m

Table 2. Description of symbols and variables.

$$\mathbf{H}_m = \sum_{i=1}^{i} \mathbf{H}_{m,i} \tag{1}$$

B_First_Review_m = Data Collection Date – First Review Date (days) (2)

$$B_{\text{Last}_{\text{Review}_{m}}} = \text{Data Collection Date} - \text{Last Review Date (days)}$$
(3)

$$B_Activity_Len_m = Last Review Date - First Review Date (days)$$
(4)

$$U_{\text{Review}}_{\text{Count}_{m}} = \frac{\sum_{n=1}^{n} U_{\text{Review}}_{\text{Count}_{m,n}}}{B_{\text{Review}}_{\text{Count}_{m}}}$$
(5)

$$U_{Fans}_{Count_{m}} = \frac{\sum_{n=1}^{n} U_{Fans}_{Count_{m,n}}}{B_{Review}_{Count_{m}}}$$
(6)

$$U_{Friends}_{Count_{m}} = \frac{\sum_{n=1}^{n} U_{Friends}_{Count_{m,n}}}{B_{Review}_{Count_{m}}}$$
(7)

$$U_Complement_Count_m = \frac{\sum_{n=1}^{n} U_Complement_Count_{m,n}}{B_Review_Count_m}$$
(8)

This study aimed to predict the \hat{H}_m for any business B_m using seven business features. These features include B_Stars_m, B_Checkin_Count_m, B_Photo_Count_m, B_Review_Count_m,

B_Activity_Len_{*m*}, B_First_Review_{*m*}, and B_Last_Review_{*m*}. In addition to business features, four reviewer features were also used to study the impact of the profile strength of the user who reviewed a business on predicting \hat{H}_m for business B_m . The pseudocode as in Algorithm 1 was used to map and generate features. The overview of cumulative helpfulness prediction is illustrated in Figure 4.



Figure 4. An overview of cumulative helpfulness prediction.

Algorithm 1: Feature Generation Pseudocode

```
Initilize B_m, for all m \in \mathcal{B};
Initilize U_n, for all n \in \mathcal{U};
Initilize R_i, for all i \in \mathcal{R};
foreach R_i written by U_n do
    Map Features R_i \leftarrow U_n;
end
foreach R<sub>i</sub> do
     Generate Features for R_i;
     Map Features B_m \leftarrow R_i;
     foreach B<sub>m</sub> do
          if R<sub>i</sub> is for B<sub>m</sub> then
           Generate Features for B<sub>m</sub>;
          else
              Skip R_i;
          end
     end
end
```

3.3. Modeling and Evaluation Metrics

Due to the numerical nature of all features, we selected LNR, GB, and NNet as learning methods in this study. The selection of learning models is also influenced by the use and performance of these models reported by previous studies. GB algorithm is an ensemble learning technique in which models are developed based on ensemble tree [56]. For the task of predicting helpfulness of large dataset of Amazon reviews, GB showed better performance than linear regression and NNet [19,36]. A NNet with three layers, namely input, hidden and output, was used for the task of helpfulness prediction showed better performance than regression methods across all datasets [38]. The linear regression did not show better performance as reported by a few studies. However, it is still the most widely adopted for the task of helpfulness prediction due to its fast execution time and explanatory power when compared with other methods [21,25,57–62]. To validate the proposed models, 10-fold cross-validation was used. The evaluation metrics used in this study were Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Squared Correlation (R^2). Total (12 (datasets) * 3 (Learning Algorithms)) * 2 (Feature Sets) = 72 models were developed and their performances were compared in this study.

4. Results and Discussion

4.1. Ranking and Comparison of Top-10 Businesses

The top k businesses ranking comparison was done using the following three criteria: (a) The ranking described in Table 3 was based on the star rating and the number of reviews. (b) The ranking based on cumulative helpfulness is given in Table 4. (c) Table 5 shows the ranking of businesses according to the star rating and cumulative helpfulness. Previously, a study classified the reviews as high-quality and low-quality reviews. In addition, high-quality reviews were further ranked based on their votes count. The difference between the actual ranking of online reviews from review portals and the predicted review ranking was used for evaluation [22]. However, in our case, there exists no list of online ranking of businesses to evaluate the results. To show the difference, we used the above-mentioned three criteria for ranking businesses.

Table 3.	Top	10	five-star	businesses	by	review	count.
					~		

Rank	Stars	Reviews	Helpfulness	Check-ins	Photos	Category	City
1	5	1936	2776	2819	95	Restaurants	Phoenix
2	5	1506	2113	13,437	0	Restaurants	Las Vegas
3	5	679	628	1456	0	Shopping	Las Vegas
4	5	662	514	211	0	Arts, Entertainment	Las Vegas
						and Events	
5	5	552	498	39	0	Home and Local	Las Vegas
						Services	
6	5	552	635	811	16	Restaurants	Mesa
7	5	543	849	2277	0	Restaurants	Las Vegas
8	5	537	2152	1852	0	Beauty and Fitness	Las Vegas
9	5	470	366	689	0	Shopping	Las Vegas
10	5	459	522	545	69	Restaurants	Henderson

Table 4. Top 10 businesses by cumulative helpfulness.

Rank	Stars	Reviews	Helpfulness	Check-ins	Photos	Category	City
1	3	546	40,635	330	0	Restaurants	Scottsdale
2	4	8339	12,701	28,872	823	Restaurants	Las Vegas
3	4	4322	9586	46,384	0	Restaurants	Las Vegas
4	3.5	6708	9337	22,225	642	Restaurants	Las Vegas
5	3.5	4206	7463	34,353	0	Travel and Hotel	Las Vegas
6	4	3055	7149	11,961	0	Nightlife	Las Vegas
7	4.5	1278	6484	3480	19	Restaurants	Scottsdale
8	3	3944	6456	30,098	0	Restaurants	Las Vegas
9	4	8348	6436	20,674	225	Restaurants	Las Vegas
10	3.5	4400	6277	9389	280	Restaurants	Las Vegas

Rank	Stars	Reviews	Helpfulness	Check-ins	Photos	Category	City
1	5	247	4105	333	4	Restaurants	Scottsdale
2	5	255	4050	150	0	Auto	Las Vegas
3	5	68	3551	84	0	Auto	Las Vegas
4	5	1936	2776	2819	95	Restaurants	Phoenix
5	5	168	2219	8	0	Home and Local	Scottsdale
						Services	
6	5	537	2152	1852	0	Beauty and Fitness	Las Vegas
7	5	1506	2113	13,437	0	Restaurants	Las Vegas
8	5	93	1864	124	0	Restaurants	Phoenix
9	5	416	1756	12	0	Home and Local	Las Vegas
						Services	Ū.
10	5	412	1391	865	0	Shopping	Las Vegas

Table 5. Top 10 5-star businesses by cumulative helpfulness.

The quality of a business is mostly judged using a star rating. However, for ranking, we cannot rely only on star rating, as many businesses have the same star rating. Similarly, a business that has a five-star rating based on 300 or more reviews cannot be treated or ranked the same as a five-star rating based on one review. To overcome this problem, a simple solution that was adopted to rank the business is by using a star rating along with review count. The Top 10 businesses ranked using this criterion are given in Table 3. In this research, we propose the idea of cumulative helpfulness as an indicator of the quality of reviews received by a business. As shown in Table 3, out the Top 10, five businesses have cumulative helpfulness even less than the number of reviews. This raises the question of whether ranking businesses using this criterion is valid. To propose a solution to this cumulative helpfulness are given in Table 4. The results seem surprising as no five-star rated businesses make it into the Top 10. Moreover, only one 4.5-star company makes it into the Top 10 using this criterion also did not seem promising, as we only considered the quality of reviews received by a business have 3-, 3.5- and 4-star ratings. Ranking using this criterion also did not seem promising, as we only considered the quality of reviews received by a business have 3-, 3.5- and ignored the quality of the business itself.

To encounter this, the ranking of business is done based on the star rating and cumulative helpfulness. The Top 10 businesses based on this ranking are given in Table 5. The results of ranking using this criterion appears to be more promising, as it ranks businesses based on rating and quality of reviews that define the rating. It is also seen that three business from the Top 10 using the first ranking criteria also make it into the Top 10 using this criterion. It is interesting to see that the second and third places are taken by businesses from the auto category, which is ignored by the previous two ranking criteria. The first place is taken by the business from restaurant category in all rankings.

4.2. Cumulative Helpfulness Prediction

We used 10-fold cross-validation method to evaluate the predictive performance of LNR, GB, and NNet for all 12 datasets using seven business features. The prediction results of cumulative helpfulness using business features were then compared with actual values to compute values for performance metrics. To see the impact of reviewer features on cumulative helpfulness, we performed the above process using seven business features along with four reviewer features. The values of RMSE, MAE and R² values for each dataset are given in Table 6. Figure 5 illustrates the feed-forward NNet model with back propagation for Nightlife dataset. The input layer takes seven business, and four reviewer features as input nodes. There are seven nodes at the hidden layer, whereas one node at the output layer. The sigmoid activation function was used in forward propagation for the hidden layer. As we were solving a regression problem, the linear activation function was used for the output layer. The weights and bias were optimized by using a back propagation algorithm to minimize the cost function. In experimental results, we used RMSE value to report the performance.



Figure 5. Multilayer neural network for nightlife.

4.3. Prediction Using Business Features

The results show that GB achieved the lowest RMSE for All category (0.534), Other (0.587) and Auto (0.643). The lowest RMSE using LNR was achieved for Restaurants (0.486), Shopping (0.605), Health (0.690), Arts, Entertainment and Events (0.412), Travel and Hotel (0.296) and Nightlife (0.345). The NNet showed the best performance on Home and Local Services (0.618), Beauty and Fitness (0.626) and Pets (0.564). LRN achieved the lowest RMSE (0.296) among all experiments on Travel and Hotel dataset. The LNR showed the best prediction performance for small- and medium-sized datasets. For large sized dataset, GB outperformed LNR and NNet. NNet also showed better performance for small and medium dataset,s as also seen in previous studies [38]. The results are in accordance with the prediction performance of GB for large datasets in comparison with other ML Models [19,36]. Overall, LNR showed the best performance over six datasets, while NNet and GB each gave the lowest RMSE for three datasets. The results show that NNet is not suitable to perform the helpfulness prediction task on large datasets.

4.4. Prediction Using Business and Reviewer Features

GB achieved the best performance for the largest dataset of All categories with the lowest RMSE of 0.530 in comparison to LNR (0.537) and NNet (0.756). In addition, GB also gave the best prediction performance for others (0.573) and Beauty and Fitness (0.619). For Restaurants (0.486), Shopping (0.608), Auto (0.623), Arts, Entertainment and Events (0.404), and Nightlife (0.337), the lowest RMSE was achieved by LNR. NNet showed the best performance for Home and Local Services (0.628), Health (0.680), Travel and Hotel (0.297) and Pets (0.595). In this experiment, GB showed the lowest RMSE for three datasets, LNR showed the best performance for five datasets and NNet gave the lowest RMSE for four datasets. The results are similar to the literature on the performance of GB over large datasets [19,36].

Detecto	ML Models	Business Features		Business + Reviewer Features			Improvement	
Datasets		RMSE	MAE	R ²	RMSE	MAE	R ²	in RMSE (%)
A 11	LNR	0.537	0.128	0.707	0.537	0.127	0.708	0%
All	GB	0.534	0.128	0.720	0.530	0.126	0.725	0.75%
categories	NNet	0.779	0.290	0.579	0.756	0.220	0.562	2.95%
	LNR	0.486	0.103	0.762	0.486	0.103	0.761	0%
Restaurants	GB	0.511	0.111	0.746	0.489	0.107	0.763	4.31%
	NNet	0.499	0.148	0.760	0.487	0.117	0.766	2.4%
	LNR	0.605	0.188	0.615	0.608	0.187	0.624	-0.5%
Shopping	GB	0.630	0.186	0.618	0.627	0.184	0.618	0.48%
	NNet	0.610	0.222	0.634	0.640	0.224	0.602	-4.92%
Home and	LNR	0.640	0.231	0.567	0.649	0.231	0.564	-1.41%
Local Services	GB	0.656	0.191	0.572	0.653	0.188	0.576	0.46%
Local Services	NNet	0.618	0.218	0.619	0.628	0.251	0.628	-1.62%
	LNR	0.591	0.213	0.652	0.594	0.210	0.643	-0.51%
Other	GB	0.587	0.215	0.665	0.573	0.205	0.679	2.39%
	NNet	0.594	0.243	0.662	0.604	0.265	0.672	-1.68%
Beauty and	LNR	0.638	0.223	0.587	0.629	0.223	0.595	1.41%
	GB	0.645	0.206	0.566	0.619	0.203	0.612	4.03%
FILLESS	NNet	0.626	0.226	0.607	0.645	0.278	0.596	-3.04%
	LNR	0.690	0.250	0.513	0.695	0.249	0.502	-0.72%
Health	GB	0.693	0.224	0.518	0.689	0.222	0.525	0.58%
	NNet	0.717	0.303	0.545	0.680	0.231	0.534	5.16%
	LNR	0.660	0.189	0.560	0.623	0.189	0.609	5.61%
Auto	GB	0.643	0.172	0.590	0.658	0.173	0.572	-2.33%
	NNet	0.704	0.294	0.612	0.648	0.214	0.611	7.95%
Arts,	LNR	0.412	0.144	0.816	0.404	0.143	0.820	1.94%
Entertainment	GB	0.451	0.149	0.788	0.457	0.148	0.786	-1.33%
and Events	NNet	0.435	0.197	0.818	0.434	0.155	0.812	0.23%
Travel and	LNR	0.296	0.098	0.882	0.301	0.098	0.886	-1.69%
Latal	GB	0.356	0.103	0.871	0.362	0.103	0.840	-1.69%
Tioter	NNet	0.326	0.151	0.874	0.297	0.122	0.910	8.9%
	LNR	0.621	0.282	0.623	0.617	0.282	0.628	0.64%
Pets	GB	0.576	0.261	0.658	0.596	0.262	0.644	-3.47%
	NNet	0.564	0.273	0.698	0.595	0.323	0.693	-5.5%
	LNR	0.345	0.131	0.872	0.337	0.130	0.844	2.32%
Nightlife	GB	0.430	0.128	0.834	0.436	0.126	0.808	-1.4%
Č.	NNet	0.490	0.173	0.811	0.452	0.157	0.784	7.76%

Table 6. Evaluation results, performance comparisons and impact of reviewer features on RMSE.

4.5. Impact of Reviewer Features on Performance

We explored the impact of reviewer features on the performance of each model by seeing percentage improvement in RMSE. For All categories, the best RMSE (0.530) was achieved by GB using business and reviewer features and showed an improvement of 0.75%. Restaurants RMSE (0.486) given by LNR remained the same using both types of feature sets. Adding reviewer features for Shopping decreased the performance, the best performance (RMSE = 0.605) being achieved using business features. The prediction performance for Home and Local Services also decreased by adding reviewer features. The performance of Other dataset increased by 2.39% using both business and reviewer features gave a boost to prediction performance for Beauty and Fitness by GB (RMSE of 0.619). For Health, the prediction performance for Auto improved with RMSE of 0.623. The performance of Arts, Entertainment, and Events improved by adding reviewer features.

For Travel and Hotel and Pets, the performance decreased by adding reviewer features. Moreover, the performance of Nightlife increased by 2.32% using business and reviewer features. The comparison of RMSE values for Nightlife and Auto are illustrated in Figures 6 and 7, respectively. Overall, out of twelve datasets, the prediction performance of one dataset showed no change, four datasets decreased and increased for seven datasets by adding reviewer features. The highest improvement in prediction was given by NNet (8.9%) for Travel and Hotel dataset by using both business and reviewer features.





Figure 6. Comparison of RMSE for Nightlife.

Figure 7. Comparison of RMSE for Auto.

The prediction performances of 36 models using business features were compared with the prediction performances of 36 models using both feature sets, as given in Table 6. The results show no change for two models, decreased performance for fifteen models and increased performance for nineteen models. We also see that, in most of the case, the model that gave the lowest performance with business features also gave the lowest performance by adding reviewer features. However, in a few cases, the best performing models also changed, e.g., GB to LNR for Auto, LNR to NNet for Travel and Hotel, LNR to NNet for Health, and NNet to GB for Beauty and Fitness. The changes in the performance of models were seen for small and medium datasets. This reflects that adding more features to small datasets can alter the performance of the model in comparison with the larger datasets.

4.6. Importance of Features

The importance of each feature related to business and reviewer varied for each dataset and model used in the experiments. However, to see the overall importance of all features in predicting the cumulative helpfulness of reviews for a business, correlation analysis was performed. Based on the correlations of features, the weights were assigned that reflect the importance of each feature. The importance of each feature is presented in Figure 8.

Among proposed business features, B_Stars was the most important feature for all datasets, as weights assigned by correlation analysis were above 0.98. The weights assigned to B_Stars were comparatively higher than other features. B_Review_Count was important for all datasets, except All Categories, Restaurant, and Nightlife. B_Checkin_Count was less significant for Restaurant and Nightlife dataset, but it appeared as an effective feature for the remaining categories. B_Photo_Count showed no importance for Pets, Auto, Health and Beauty, and Fitness, but was important for the remaining datasets. The B_Activity_Len and B_First_Review appeared to be the most effective features for Beauty and Fitness, Health and Auto, compared to their importance for the remaining datasets. The importance of these features has also been reported by previous studies in helpfulness prediction of reviews [36]. Lastly, B_Last_Review appeared to be the most important feature for all datasets.

When looking at the reviewer features, we found that U_Review_Count, U_Friends_Count, and U_Compliment_Count were more significant for Pets, Auto, Health and Beauty and Fitness in comparison to their importance for the other datasets. U_Fan was an effective feature for Pets, Auto, Health and Beauty and Fitness, whereas it showed no importance for the other features. Overall, B_Star, B_Review_Count, and B_Last_Review appeared to be the most important features among all datasets.



Figure 8. The importance of each feature for all datasets.

5. Implications

This study has both theoretical as well as practical implications. From a theoretical perspective, the cumulative helpfulness for a business proposed and its use in ranking businesses will encourage researchers and academics to explore further the problem of predicting helpfulness from this perspective. Previously, researchers ranked online reviews and reviewers [22]. However, this study paves a new way of ranking online products, businesses, and services by combining both average star rating and cumulative helpfulness of a business. The best prediction results were achieved by the

combination of business and reviewer features for the majority of datasets. However, there are few datasets in which the use of both business and reviewer features reduced the prediction performance. The results of this study will encourage researchers to further explore and verify the proposed features on different datasets. This study also shows the important features for predicting cumulative helpfulness that can be used in future studies. The experimental results validated that GB shows the best performance for the large dataset. In short, this study created a whole new dimension to investigate the problem of helpfulness prediction from a business perspective that was previously ignored. Moreover, it will be interesting to study the impact of cumulative helpfulness of reviews for business on review helpfulness prediction and ranking of reviews and reviewers.

The practical implications of this study include a new criterion to rank the businesses that will be more helpful to the user in identifying the business with more quality review. Previously, the information on most of the review platforms is the average star rating and review count. The comparison of ranking strategy will encourage the review platforms to make the stats related to business more useful to viewers by adding cumulative helpfulness score. To further ease the users in searching and selecting businesses, the proposed ranking criteria can be used for ranking businesses or products based on individual category and location. This study gives insights to companies in exploring the importance of factors to make strategic changes and controlling features that will bring more quality reviews for a business. Moreover, the experimental results and performance of different ML algorithms for different sizes of datasets will also guide the practitioners in selecting an appropriate learning algorithm.

6. Limitations and Future Work

As with other studies, this study also has several limitations. Firstly, the cumulative helpfulness is used for ranking and prediction only for a dataset of businesses from Yelp.com. The future work will consider the application of the proposed features for businesses, products, and services from other popular review platforms and e-commerce websites. Secondly, to evaluate the review ranking, the previous studies matched it with actual online reviews of review portals. However, in this study, there exists no online ranking of business with which it can be compared. A future extension of this work will be focused entirely on the ranking of businesses where a valid measuring metric will be proposed to verify the ranking of businesses. In addition, it will be compared with the existing ranking techniques for reviewers and reviews. Thirdly, this study illustrates the use of cumulative helpfulness for businesses for all categories. However, future studies can rank a business based on the individual category and location. Fourthly, only numerical features of reviews and reviewers were used in this study. The impact of textural features of reviews received by a business in predicting cumulative helpfulness should be explored. Lastly, the use of cumulative helpfulness of reviews for a business for the ranking purpose will attract fake votes, similar to the fake reviews that affect the overall ranking. Therefore, future research should also take into consideration the detection of fake votes, along with the detection of fake reviews. The researchers can further explore and enhance the performance of predicting cumulative helpfulness by identifying and testing new features using DL models. The future studies can use cumulative helpfulness of reviews for a business along with other features to perform prediction and ranking tasks associated with crowd-sourced reviews.

7. Conclusions

The increased volume of reviews makes it difficult for consumers and retailers to evaluate the quality of products. The importance of reviews has encouraged researchers to model and predict the helpfulness of the reviews as it becomes a critical factor for consumers in making purchase decisions. However, helpfulness for "review" and "reviewer" has only been focused on by previous studies. This study proposed the concept of cumulative helpfulness of reviews for a "business" that make it easy for consumers and business managers to see the overall quality of reviews received by a business. The applicability of cumulative helpfulness in raking businesses was illustrated by using a real-time

dataset of 1,92,606 businesses from Yelp.com. The ranking of business using star rating along with cumulative helpfulness appears more reliable.

Further, the prediction of cumulative helpfulness is performed using seven business and four reviewer features. The prediction performance for LNR, GB and NNet were compared on datasets of different size and categories. GB outperformed LNR and NNet for a large dataset of all categories, however, for small and medium datasets, LNR and NNet performed better than GB. The use of reviewer feature, along with business features, shows a significant improvement in performance for predicting cumulative helpfulness. When examining the individual importance of features, business star rating, review count and the number of days since the last review appear to be the most important features. This study will help customers and retailers to see the overall quality of reviews for a business. The review platform can rank firms in a better way using their cumulative helpfulness score. Future work can evaluate business ranking, validate the impact of proposed features for predicting helpfulness and ranking tasks and examine performance comparison with deep learning techniques.

Author Contributions: All authors have the same contribution in the manuscript.

Funding: This work was supported under the framework of international cooperation program managed by the National Research Foundation of Korea (2018090561 and FY2018). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017030223).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LNR	Linear Regression
GB	Gradient Boosting
NNet	Neural Network
e-WoM	electronic Word-of-Mouth
UGC	User-Generated Content
ML	Machine Learning
RandF	Random Forest
LGR	Logistic Regression
SVR	Support Vector Regression
M5P	Model tree
DNN	Deep Neural Network
DT	Decision Tree
SVM	Support Vector Machine
CNN	Convolutional Neural Network
RFM	Recency, Frequency, and Monetary
DL	Deep Learning
NLP	Natural Language Processing
NB	Naïve Bayes
BoW	Bag-of-Words

References

- 1. Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng.* **2010**, *23*, 1498–1512. [CrossRef]
- 2. Cheung, C.M.; Thadani, D.R. The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decis. Support Syst.* **2012**, *54*, 461–470. [CrossRef]
- Zhao, Y.; Yang, S.; Narayan, V.; Zhao, Y. Modeling consumer learning from online product reviews. *Mark. Sci.* 2013, 32, 153–169. [CrossRef]
- 4. Zhu, F.; Zhang, X. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Mark.* **2010**, *74*, 133–148. [CrossRef]

- Krishnamoorthy, S. Linguistic features for review helpfulness prediction. *Expert Syst. Appl.* 2015, 42, 3751–3759. [CrossRef]
- 6. Tsao, W.C.; Hsieh, M.T. eWOM persuasiveness: Do eWOM platforms and product type matter? *Electron. Commer. Res.* **2015**, *15*, 509–541. [CrossRef]
- Diaz, G.O.; Ng, V. Modeling and Prediction of Online Product Review Helpfulness: A Survey. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 698–708.
- 8. BrightLocal. Local Consumer Review Survey. Online Reviews Statistics & Trends. 2019. Available online: http://www.brightlocal.com/learn/local-consumer-review-survey/ (accessed on 27 June 2019).
- 9. TripAdvisor. Media Centre. Available online: https://tripadvisor.mediaroom.com/uk-about-us (accessed on 27 June 2019).
- 10. Yelp. About Us. Available online: https://www.yelp.co.uk/about (accessed on 27 June 2019).
- 11. Berger, J. Word of mouth and interpersonal communication: A review and directions for future research. *J. Consum. Psychol.* **2014**, *24*, 586–607. [CrossRef]
- Dirsehan, T. An application of text mining to capture and analyze eWOM: A pilot study on tourism sector. In *Capturing, Analyzing, and Managing Word-of-Mouth in the Digital Marketplace*; IGI Global: Hershey, PA, USA, 2016; pp. 168–186.
- 13. Bilal, M.; Gani, A.; Ullah Lali, M.I.; Marjani, M.; Malik, N. Social Profiling: A Review, Taxonomy, and Challenges. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 433–450. [CrossRef]
- 14. Chen, J.; Chen, Y.; Du, X.; Li, C.; Lu, J.; Zhao, S.; Zhou, X. Big data challenge: A data management perspective. *Front. Comput. Sci.* **2013**, *7*, 157–164. [CrossRef]
- 15. Cao, Q.; Duan, W.; Gan, Q. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decis. Support Syst.* **2011**, *50*, 511–521. [CrossRef]
- 16. Li, M.; Huang, L.; Tan, C.H.; Wei, K.K. Helpfulness of online product reviews as seen by consumers: Source and content features. *Int. J. Electron. Commer.* **2013**, *17*, 101–136. [CrossRef]
- 17. Liu, Y.; Huang, X.; An, A.; Yu, X. Modeling and predicting the helpfulness of online reviews. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 443–452.
- Furner, C.P.; Zinko, R.A. The influence of information overload on the development of trust and purchase intention based on online product reviews in a mobile vs. web environment: An empirical investigation. *Electron. Mark.* 2017, 27, 211–224. [CrossRef]
- 19. Singh, J.P.; Irani, S.; Rana, N.P.; Dwivedi, Y.K.; Saumya, S.; Roy, P.K. Predicting the "helpfulness" of online consumer reviews. *J. Bus. Res.* 2017, *70*, 346–355. [CrossRef]
- 20. Hong, H.; Xu, D.; Wang, G.A.; Fan, W. Understanding the determinants of online review helpfulness: A meta-analytic investigation. *Decis. Support Syst.* **2017**, *102*, 1–11. [CrossRef]
- 21. Zhang, Y.; Lin, Z. Predicting the helpfulness of online product reviews: A multilingual approach. *Electron. Commer. Res. Appl.* **2018**, 27, 1–10. [CrossRef]
- 22. Saumya, S.; Singh, J.P.; Baabdullah, A.M.; Rana, N.P.; Dwivedi, Y.K. Ranking online consumer reviews. *Electron. Commer. Res. Appl.* **2018**, 29, 78–89. [CrossRef]
- 23. Saumya, S.; Singh, J.P.; Dwivedi, Y.K. Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Comput.* **2019**, 1–17. [CrossRef]
- 24. Arif, M.; Qamar, U.; Khan, F.H.; Bashir, S. A Survey of Customer Review Helpfulness Prediction Techniques. In *Proceedings of SAI Intelligent Systems Conference*; Springer: Cham, Switzerland, 2018; pp. 215–226.
- Mudambi, S.M.; Schuff, D. What makes a helpful review? A study of customer reviews on Amazon.com. MIS Q. 2010, 34, 185–200. [CrossRef]
- 26. Salehan, M.; Kim, D.J. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decis. Support Syst.* **2016**, *81*, 30–40. [CrossRef]
- 27. Eslami, S.P.; Ghasemaghaei, M.; Hassanein, K. Which online reviews do consumers find most helpful? A multi-method investigation. *Decis. Support Syst.* **2018**, *113*, 32–42. [CrossRef]
- Li, X.; Hitt, L.M. Price effects in online product reviews: An analytical model and empirical analysis. *MIS Q.* 2010, 34, 809–831. [CrossRef]
- 29. Anderson, M.; Magruder, J. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *Econ. J.* **2012**, *122*, 957–989. [CrossRef]

- 30. Yan, X.; Wang, J.; Chau, M. Customer revisit intention to restaurants: Evidence from online reviews. *Inf. Syst. Front.* **2015**, 17, 645–657. [CrossRef]
- 31. Spool, J.M. The Magic Behind Amazon's 2.7 Billion Dollar Question. 2016. Available online: https://articles.uie.com/magicbehindamazon/ (accessed on 27 June 2019).
- Lu, Y.; Tsaparas, P.; Ntoulas, A.; Polanyi, L. Exploiting social context for review quality prediction. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 691–700.
- Kim, S.M.; Pantel, P.; Chklovski, T.; Pennacchiotti, M. Automatically assessing review helpfulness. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 22–23 July 2006; pp. 423–430.
- 34. Tang, J.; Gao, H.; Hu, X.; Liu, H. Context-aware review helpfulness rating prediction. In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; pp. 1–8.
- 35. Hu, Y.H.; Chen, K. Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *Int. J. Inf. Manag.* **2016**, *36*, 929–944. [CrossRef]
- 36. Malik, M.; Hussain, A. An analysis of review content and reviewer variables that contribute to review helpfulness. *Inf. Process. Manag.* **2018**, *54*, 88–104. [CrossRef]
- 37. Huang, A.H.; Chen, K.; Yen, D.C.; Tran, T.P. A study of factors that contribute to online review helpfulness. *Comput. Hum. Behav.* **2015**, *48*, 17–27. [CrossRef]
- 38. Lee, S.; Choeh, J.Y. Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Syst. Appl.* **2014**, *41*, 3041–3046. [CrossRef]
- Wu, J. Review popularity and review helpfulness: A model for user review effectiveness. *Decis. Support Syst.* 2017, 97, 92–103. [CrossRef]
- 40. Malik, M.; Hussain, A. Exploring the influential reviewer, review and product determinants for review helpfulness. *Artif. Intell. Rev.* **2018**, 1–21. [CrossRef]
- 41. Malik, M.; Hussain, A. Helpfulness of product reviews as a function of discrete positive and negative emotions. *Comput. Hum. Behav.* **2017**, *73*, 290–302. [CrossRef]
- 42. Akbarabadi, M.; Hosseini, M. Predicting the helpfulness of online customer reviews: The role of title features. *Int. J. Mark. Res.* **2018**. [CrossRef]
- 43. Gao, B.; Hu, N.; Bose, I. Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decis. Support Syst.* **2017**, *95*, 1–11. [CrossRef]
- 44. Qazi, A.; Syed, K.B.S.; Raj, R.G.; Cambria, E.; Tahir, M.; Alghazzawi, D. A concept-level approach to the analysis of online review helpfulness. *Comput. Hum. Behav.* **2016**, *58*, 75–81. [CrossRef]
- 45. Lee, P.J.; Hu, Y.H.; Lu, K.T. Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telemat. Inform.* **2018**, *35*, 436–445. [CrossRef]
- 46. Zhou, Y.; Yang, S. Roles of review numerical and textual characteristics on review helpfulness across three different types of reviews. *IEEE Access* **2019**, *7*, 27769–27780. [CrossRef]
- 47. Hu, Y.H.; Chen, K.; Lee, P.J. The effect of user-controllable filters on the prediction of online hotel reviews. *Inf. Manag.* 2017, 54, 728–744. [CrossRef]
- 48. Ngo-Ye, T.L.; Sinha, A.P. The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decis. Support Syst.* **2014**, *61*, 47–58. [CrossRef]
- 49. Mohammadiani, R.P.; Mohammadi, S.; Malik, Z. Understanding the relationship strengths in users' activities, review helpfulness and influence. *Comput. Hum. Behav.* **2017**, *75*, 117–129. [CrossRef]
- 50. Ma, Y.; Xiang, Z.; Du, Q.; Fan, W. Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *Int. J. Hosp. Manag.* **2018**, *71*, 120–131. [CrossRef]
- Karimi, S.; Wang, F. Online review helpfulness: Impact of reviewer profile image. *Decis. Support Syst.* 2017, 96, 39–48. [CrossRef]
- 52. Shin, S.; Du, Q.; Xiang, Z. What's Vs. How's in Online Hotel Reviews: Comparing Information Value of Content and Writing Style with Machine Learning. In *Information and Communication Technologies in Tourism 2019*; Springer: Berlin, Germany, 2019; pp. 321–332.
- 53. Li, S.T.; Pham, T.T.; Chuang, H.C. Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles. *Inf. Manag.* **2019**, *56*, 28–38. [CrossRef]
- 54. Ngo-Ye, T.L.; Sinha, A.P.; Sen, A. Predicting the helpfulness of online reviews using a scripts-enriched text regression model. *Expert Syst. Appl.* **2017**, *71*, 98–110. [CrossRef]

- 55. Yelp. Yelp Dataset Challenge. Available online: https://www.yelp.com/dataset/challenge (accessed on 27 June 2019).
- 56. Friedman, J.H. Stochastic gradient boosting. Comput. Stat. Data Anal. 2002, 38, 367–378. [CrossRef]
- 57. Yin, D.; Bond, S.; Zhang, H. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Q.* **2014**, *38*, 539–560. [CrossRef]
- 58. Yang, Y.; Yan, Y.; Qiu, M.; Bao, F. Semantic analysis and helpfulness prediction of text for online product reviews. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; Volume 2, pp. 38–44.
- Korfiatis, N.; GarcíA-Bariocanal, E.; SáNchez-Alonso, S. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electron. Commer. Res. Appl.* 2012, 11, 205–217. [CrossRef]
- 60. Forman, C.; Ghose, A.; Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inf. Syst. Res.* **2008**, *19*, 291–313. [CrossRef]
- 61. Chevalier, J.A.; Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *J. Mark. Res.* **2006**, *43*, 345–354. [CrossRef]
- 62. Otterbacher, J. 'Helpfulness' in online communities: A measure of message quality. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 4–9 April 2009; pp. 955–964.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).