

## Review Helpfulness Prediction: Survey

Yasamyian Almutairi<sup>1</sup>, Manal Abdullah<sup>2</sup>, Dimah Alahmadi<sup>3</sup>

<sup>1</sup> Master student in Information system at king Abdulaziz university

<sup>2</sup> Associated Professor in king Abdulaziz university

<sup>3</sup> Assistant Professor in king Abdulaziz university

---

### Article Info

Received Dec 28<sup>th</sup>, 2018

---

#### Keyword:

Review helpfulness,  
Recommender system

---

### ABSTRACT

Online reviews have become the major driving factor influencing purchasing behavior and patterns of social customers. However, it is difficult for customer to cover good reviews about any product or service according to massive amount of reviews latest years. Many previous researches provide innovative models about predicting review helpfulness in E-commerce websites. Some of these studies exploring the direct effect of review attributes on review helpfulness while others focused on reviewer's attributes only. The main objective of this research is to review the most important attributes that have an affect on review helpfulness from many perspectives such as datasets, techniques, frameworks and evaluation methods of the experiments. The paper ends up with important findings about most attributes effect the review helpfulness such as Review Valence.

---

#### Corresponding Author:

Second Author,  
Departement of Information System,  
King Abdulaziz University,  
Jeddah, Saudi Arabia,

Email: [maaabdullah@kau.edu.sa](mailto:maaabdullah@kau.edu.sa)

---

### 1. Introduction

Online reviews are become important factor in assisting customers' buying decisions. Reviews offer valuable information can influence customers' opinion. Moore in [1] states that 92% of customers nowadays read online reviews. This makes online review helpfulness more key factor in E-commerce platform. Moreover, online reviews differ in their support to customers due to different subjective. For example, some of customers discover the reviews that support their decision-making such as product evaluation reviews. In other words, they looks for its utility that called "review helpfulness" [2]. Review helpfulness indicates whether the review gives useful product assessment and buying decision to other customers. Hence, it is important to explore attributes and which make the review more helpful. These attributes belong to two major categories: the first is about the review itself such as review length, rating valence, and review extremity [3]. The second is related to reviewers such as reviewer Social Profile Information (SPI), reviewer ranking, and reviewer engagement [4].

Due to the importance of the helpfulness of reviews to the social customers and the numerous profits to the E-commerce website, this research paper discusses the main perspectives of reviews helpfulness prediction.

In next section, we will identify the concept of review helpfulness in E-commerce websites then illustrate the important attributes that affect the review helpfulness. The next sections are revised of dataset, techniques, frameworks and evaluation methods of the review helpfulness prediction.

## 2. Review Helpfulness

Review “helpfulness” measures whether a review gives product assessment and buying decision to customers [5]. In other words, its effectiveness is not equal in their value to consumers. Customers are more interested to reviews seeming more helpful to them. Therefore, websites that categorize the helpful reviews gain higher consumer attention [6].

In a voting system as the one offered by Amazon [7] the review helpfulness can be defined as:

$$H = \frac{np}{np+nn} \quad (1)$$

Where  $np$  is the number of positive votes and  $nn$  is the number of negative votes [8]. While this method is easy and good enough it still show some shortcoming, such as the lack of votes for fresh reviews [9] and the point that not all readers votes on the review [8]. Therefore, the most-voted reviews are not certainly reflecting the exact representation of the most helpful ones.

## 3. Attributes affects Review Helpfulness

Many attributes can Increase the power of the review helpfulness. Some of these attributes are related to review itself and others are related to reviewer who writes this review. In this context, we will emphasize on the most important attributes have been confirmed to have impact on review helpfulness.

shows all perspectives that we will highlights in this review paper in context of review helpfulness. We will start with review attributes then we move to the second perspective that focus on reviewers' attributes.

### 3.1 Review Attributes

Review attributes display all criteria that are related to written text of the review. The attributes may differ from platform to another according to nature of this platform. In addition, review attributes can affect review helpfulness in different percentage due to the power impact of this attribute.

summarize the most important review attributes.

First study by Salehan, and Kim in [4] focused on the following variables: Title length, Review sentiment, Title sentiment, Longevity, Review length, Review polarity and Title Polarity. They found that negative sentiment influences the performance of online reviews.

Other study by Jianan Wu in [10] examined three determinants as following:

1. Review attributes (Valence, Depth, and Life),
2. Reviewer attribute (Credibility),
3. Review hosting website attribute (Framing).

In addition, Barbosa Jardeson LN and Moura Raimundo Santos in [11] hypothesized that there are many important attributes influence review helpfulness. One of them is textual attributes, which include text readability, text size and sentiment expressed in the text. Also, Review meta-data attribute which is the final evaluation (“recommended” or “not recommended”) and posting date of the review.

Yadong Zhang in [12] experimented the helpfulness of reviews according to the following variables: Consumer rating, Name, Product review, Date, Reviewers' ranking, Reviewers' helpful vote percentage, Review number, Helpful vote and Total vote.

In addition, Ahmed Erfan In [13] and Singh Williamjeet in [14] concentrated on sentence stage analysis and split the sentence in two classes (positive and negative). Then they observed the effect of sentiment attribute on review helpfulness.

While Gao Baojun and Hu Nan in [15] aimed to examine whether attributes can be used to predict the helpfulness of future reviews. They addressed review length, ratings, and equivocality that proved to have significant effect on review helpfulness.

On other hand, Ghose Anindya and Ipeirotis Panagiotis in [7] explored how the review and the self-reported characteristics of the reviewer can effect online community and social behavior such as review helpfulness. To examine this, they collected data from Amazon.com and analyzed the associated review system. For each review, they retrieved the real written content of the review and the review rating of the product assumed by the reviewer. The textual analysis of reviews includes Readability and Subjectivity.

Another study by Lu Shuya and Wu Jianan In [16] used different strategies by add three time measures called *dynamic drivers of a review's helpfulness*:

1. Observation day  $d$  (number of days spend from product distribution and the observed date).
2. Publish and post lifetime  $L$  (number of days spend since the review is published).
3. Timing  $T$  (number of days spend between the day the product release and the day the review is published)

Also, they used another measure called *static drivers of a review's helpfulness* includes quality attributes of a review such as valence, trustworthiness and Length.

In addition, Hong Hong and Xu Di in [17] classified the attributes of review helpfulness into two groups: one of them Review-related attributes such as (review depth, review readability, linear review rating, quadratic review rating, and review age).

Wang Xi and Tang Liang Rebecca in [18] used 8 emotional measurements and Linguistic style matching (LSM). The emotional vocabularies and useful vocabularies for LSM were detected and examined using the Linguistic Inquiry and Word Count (LIWC) 2015. LIWC offer helpful process to analyze individuals' speaking and written speech samples [19]. LIWC estimate the ratio of vocabularies in a group compared with its private dictionaries. In addition, it generates scores of each eight emotional measurements and functional words.

Moreover, Kaushik, Kapil Mishra, Rajhans in [20] addressed 7 attributes:

1. Percentage of confirmed purchase reviews,
2. Helpfulness amount of helpful reviews,
3. Balance of helpful reviews,
4. Sequence of negative reviews in Helpful review set,
5. Amount of helpful reviews showed on product page,
6. Average informative of helpful reviews showed on product page.
7. Arrangement of Informative among helpful review set.

In addition, they used other five control variables (Savings, Price, Product age, Number of Reviews on product page, Average Rating of Product) and one dependent variable (Sale).

Finally, ChenYuanlin and Chai Yueting in [21] explored the influence of some review attributes such as review readers, ratings and number of votes. Beside other reviewers and platform attributes.

In conclusion, the most papers proved the strong relationship between some review attributes and review helpfulness. The most important attributes that have strong relationship are review length, review ratings, review valence, review trustworthiness, review depth, review life and review sentiment.

To sum up, all review attributes mention above have at least 10% impact on review helpfulness. However, according to the main discoveries of this paper, we recommend to focus on the most three attributes that have greatest impact on review helpfulness. The three attributes are: review sentiment, review ratings and review life or age. Many of future investigation can explore the characteristics of these attributes that make it the most important attributes have impact on review helpfulness.

### 3.2 Reviewers' Attributes

Reviewers' attributes are varying from E-commerce platform to another according to information provided in reviewer profile platform. The most common reviewer's attributes that affect review helpfulness will be reviewed in the following.

First, Siering Michael and Muntermann Jan in [22] investigated the impact of reviewer-related attributes such as reviewer expertise and reviewer non-anonymity on review helpfulness. Furthermore, they consider other control variables include Review depth, review readability, and review extremity as content-related attributes.

Moreover, Hong Hong, Xu Di and Wang G Alan in [17] classified the attributes of review helpfulness into two groups one of them Reviewer-related factors such as (Reviewer information disclosure, Reviewer expertise, Reviewer expert label, Number of reviewer friends and fans).

Also, ChenYuanlin and Chai Yueting in [21] explored the influence of some reviewer and review attributes on review helpfulness. They focused on the helpfulness of reviews from the customers' perspective such reviewer characteristic, experience, expert, personal subjective feelings and preference. They also addressed the impact of review readers, ratings and number of votes.

Another study by Ghose Anindya and Ipeirotis Panagiotis in [7] explored how the review and the self-reported characteristics of the reviewer can affect online community and social behavior such as review helpfulness. They focused on the following attributes:

1. **Reviewer Disclosure:** whether the user disclosed their real name, their location, nickname, and

hobbies reviewer rank.

2. **Reviewer History:** they collected the past reviews for every reviewer and gathered the helpful and total votes for every past reviews.

On other hand, Barbosa Jardeson LN and Moura Raimundo Santos in [11] hypothesized that there are two important attributes influence review helpfulness. These attributes are based on the authorship attributes include Author Reputation such as (Average number of reviews votes, Average number of positive votes, Number of friends a client has in the Steam community). Author Expertise attributes contains number of hours a review author played the analyzed game.

On other hand, Gao Baojun and Hu Nan in [15] aimed to examine whether reviewers attributes can be used to predict the helpfulness of future reviews. They studied some reviewer attributes such as reviewer's ID, name, age, gender, location, number of cities visited, entire amount of reviews, ratings, and helpful votes (considering the reviewer's experience).

While Ngo-Ye Thomas and Sinha Atish in [23] proposed new reviewers' attributes analysis called RFM. RFM means Recency, Frequency, and Monetary Value of a reviewer history. The researcher calculated RFM variables by: **Recency** the amount of days between a reviewer's current review and the immediate previous one. **Frequency**: the amount of reviews a reviewer has created and published on the website before the current review. **Monetary Value**: the average number of helpfulness votes obtained by the reviewer through all the reviews he or she has written.

In addition, Goswami Kunal, Park Younghee and Song Chungsik in [24] used some attributes that indicate reviewers' social interaction behavior. They considered all attributes that show the level of activity of the reviewers in a website. They called this attributes Reviewer's digital personality include: Number of friends, Number of votes, Number of compliments, Number of followers and Photo count.

Finally, Karimi Sahar and Wang Fangn in [25] analyzed the impact of reviewer image on review helpfulness. They hypothesized that "if reviewer profile image available, it has a positive effect on review helpfulness". Beside the reviewer image, they suggest that it may interact with review attributes such as review length, rating valence, and review equivocality to resulting stronger effect on review helpfulness. The most important attributes are visualized in

and

Table 1 summarizes the attributes and their references.

#### 4. Datasets used in review helpfulness prediction

The experiments in the literature use diverse reviews datasets collected from different platforms see . For example the dataset of [13] was collected from aclimdb. They have 22000 movie reviews for train set and 3001 movie reviews for test set.

While the experiment in [26] used 200 of Arabic user reviews from Forums, Facebook, YouTube, and google search. The researchers stored the reviews in a database tables, then they started data processing by cleaning the text and associated the ratings with the reviews then, they annotated the corpus to organize it into one format (positive, negative, or neutral).

The experiment in [4] done by collect 35,000 online reviews of 20 diverse products from Amazon.com by applying crawler software established by the authors. The final sample consisted of 2616 reviews.

Another study in [10] examined the effects of some attributes on review helpfulness by collected reviews of two products at Amazon: TV streaming box and PlayStation 3. The number of reviews was 9484 for first product and 3272 for the second one.

On other hand, the researcher in [11] collected reviews from Steam witch is online community that enables users to have profile and add a personal photo, have friends, get private and public messages, publish screenshots, texts and videos and involve in groups.

The data gathering for the model training and testing done by using a scraper written in Python. They focused only in the 100 most played games on Steam. The preliminary dataset had 10,168 reviews after the data cleaning; they obtained a final dataset containing 5,823 reviews.

The researchers in [23] collected data from Amazon book reviews and Yelp restaurant reviews. As a result, they had a group of 584 Amazon book reviews and 7465 Yelp restaurant reviews.

Another research in [24] collected data from Yelp.com using a web crawler. They took a subset of this dataset that consists of 135,413 reviews and 66,936 unique reviewers.

Other experiment in [14] collected 3900 different mobile reviews from various website.

Also, The experiment of [27] used different category of dataset to analyze the performance of their Lexicon based classifier. They collected reviews manually from various Ecommerce websites such as Snapdeal, Amazon and FlipKart.also they use CrowedFlower’s text\_emotion dataset.

The study in [15] merged different datasets from TripAdvisor.com for the experimental study. The first step, they collect all New York City hotel reviews, including review data, reviewer data and hotel data. After data cleaning and pre-processing they end up with 8676 reviews

Another study in [25] randomly collected 2178 reviews of nine mobile gaming apps from Google Play.

**Table 1 . Review and Reviewer attributes.**

Attributes field	Attributes name	References
Review attributes	Review readability	[10] [12] [7] [22]
	Review and Title sentiment	[10] [4] [13] [14] [20]
	Review Depth	[10] [12] [22]
	Review and Title length	[4] [12] [25] [16] [28]
	Review polarity	[4] [12]
	Review Life or age	[10] [21] [17]
	Review Subjectivity	[7]
	Review ratings	[12] [15] [21] [17] [20]
	Review extremity	[22]
	Review trustworthiness	[16]
	Observation day	[21]
	Number of votes	[17] [20]
	Reviewer attributes	Reviewer Credibility
Reviewer Reputation		[11]
Reviewer Expertise		[11] [15]
Reviewer Disclosure		[7]
Reviewer History		[23] [7]
Reviewer Recency		[23]
Reviewer Monetary		[23]
Reviewer Frequency		[23]
Reviewer’s digital personality		[24] [17]
Reviewer image		[25]
Reviewer expert label	[17]	
Platform attributes	Framing	[10]

Also, the study in [7] done by collected data from Amazon.com and examined the related review system. They collected all reviews of three product (DVD, audio & video and Digital camera) since the product was released into the website.

Moreover, The paper [21] used review dataset gathered from Amazon by McAuley and Leskovec [29]. This dataset included 34 686 770 reviews, 6 643 669 consumers, and 2 441 053 products. Each review in dataset were created between 1995 and 2013. In addition, the researchers chose top 100 reviews who reviewed the most times.

The papers [28] collected a huge numbers of reviews and reviews relevant information, such as review contents, reviewer information and product details . The dataset offered by Yelp.com, it is collected between October 2004 and December 2015. The final dataset contained reviews from 10 cities of four countries. They

use four different sources to categorize the variables into review, restaurant, reviewer, and consumer check-in information.

For the experiment in [16] the authors used Helium Scraper crawler software to collect data from Amazon.com. They chose one product and gathered reviews every day since the product release date. The dataset included 29,426 reviews/day for 40-day.

On other hand, the experiment in [17] used scientific studies as dataset. They used different ways to search the literature. For English studies, they used papers from most common digital library databases such as Science Direct, EBSCO, SAGE, and Taylor & Francis. In addition, they explored some papers manually from five-respected information systems (IS) journals to observed review helpfulness are most likely published. The final dataset for meta-analysis contained forty-two papers according to [30] suggestion that states a meta-analysis should have at least fifteen paper.

In [18] Data was gathered from Yelp during October 10–16, 2017 from top 10 travel destinations in the U.S. they chose top 10 restaurants in each destination. The final dataset end up with 265,205 reviews. Furthermore, they collected additional information including consumer elite rank, review elapsed date, reviewer rating of the restaurant and restaurant category.

In addition, the data in [20] collected from amazon.in platform for many products. They picked a product that has a useful participation such as electronics products. First, they chose 100 mobiles initiated in past three months from 1 January 2016. They crawled data for nine months every week by using R Week-by-Week. Then, they obtained important variables by using XML package in R.

Finally, in [22] the dataset collected from Amazon's website for two product types. They collected data from different product kinds and picked 100 best-selling products.

précis the most used datasets in the literature and

Table 2 summarize the datasets with their references.

**Table 2: Datasets used in review helpfulness prediction**

Dataset	References
Amazon	[4] [10] [23] [7] [21] [16] [20] [22] [27]
Yelp	[23] [24] [18] [28]
TripAdvisor	[15]
Steam	[11]
Snapdeal, FlipKart.	[27]
Digital library databases	[17]
Google Play	[25]

### 5. Techniques used in review helpfulness prediction

As demonstrated in

many techniques of review helpfulness prediction are used in the literature.

First, the experiment in [4] used some regression equations to prove the correlation between these attributes and review helpfulness. On the other hand, the researcher in [11] used ANN MLP for prediction and the Correlation-based Feature Selection (CFS) to validation.

While the experiment in [12] produced the predictive model by the SVM, one of the top supervised machine learning algorithms. They applied the algorithms by using song WEKA tool [31].

Also, The researchers in [23] designed a text regression test to predict review helpfulness by using support vector regression (SVR).

Another research in [24] experimented various combinations of the attributes that affect the review helpfulness with the neural network algorithm.

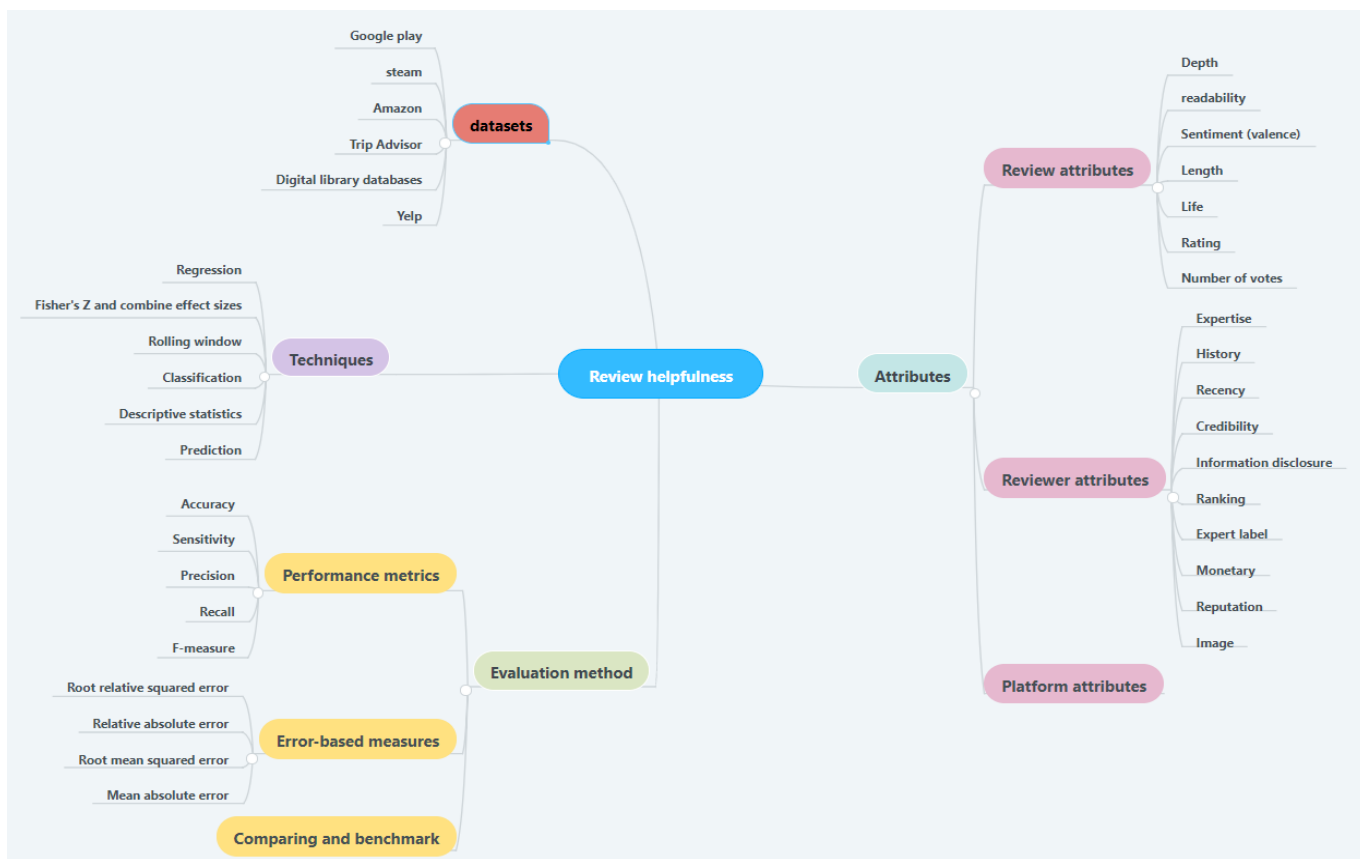
And the experiment in [13] implemented using SVM algorithm for sentiment analysis with two different kernels: Poly Kernel and RBF Kernel.

#### Figure 1: Paper classification of review helpfulness.

Also, Authors repeated the experiment with different algorithms such as MLP and Naïve Bayes. The experiment in [14] was done by implement some code in MATLAB. The researcher collected positive and negative reviews based on the normal features of Camera, Battery, Screen, Sound, Design and hardware/software Performing.

The study in domain of hotels [15] prepared by regressed rating deviation on reviewers' absolute bias in rating controlling for the reviewer, review, and hotel attributes.

Another study [25] applied regression and descriptive statistics, including mean and standard deviation for all variables and their Pearson correlation coefficients.



In addition, the study [7] decided to use two approaches regression and classification approach to build a dual prediction model that categorizes a review as helpful or not. They did the classification experiment with Support Vector Machines and Random Forests.

In [21] the authors used 2 different parts of statistical tests. The first part is building a Mann-Kendall test on 100 valance helpfulness vectors. The second part is building an accumulative Mann-Kendall test on 100\_5 helpfulness-valance matrix.

On other hand , in [16] the researcher used a rolling window technique to assess the impacts through post lifespan L and post timing T in a series of data time windows (DTWs). Therefore, they calculated Models 1–3 using a panel vector auto-regression (PVAR) estimation for every DTW. To do this estimation, they performed an advance orthogonal deviation transformation [32] to address panel specification.

In [17] the experiment done by using Fisher's Z and link effect sizes (i.e., correlation coefficients) by using Comprehensive Meta-Analysis (CMA) 2.0 software.

While the authors in [20] used linear regression to test the correlations in proposed model. They ran the regression among the content of product page and following week sales of the product. In this study, they have used eight input variables and five independent variables (Savings, Price, Product age, Number of Reviews on product page, Average Rating of Product) and one dependent variable (Sale).

Finally in [22] the researcher applied a Tobit regression analysis to observe the impact of different attributes on review helpfulness [33]. The Table 3 illustrate the most techniques used in the literature.

**Table 3: Techniques used in review helpfulness prediction.**

Techniques field	Techniques type	References
<b>Regression</b>	Tobit	[16] [22]
	linear regression	[20]
	support vector regression (SVR)	[23]
<b>Fisher's Z and combine effect sizes</b>	correlation coefficients	[25] [17]
<b>rolling window</b>	panel vector auto-regression (PVAR)	[16]
<b>Classification</b>	Support Vector Machines	[7]
	Random Forests	[7]
<b>descriptive statistics</b>	Mean, standard deviation and Pearson correlation coefficients	[25]
	Mann-Kendall	[21]
<b>Prediction</b>	ANN MLP	[11]
	MATLAB code	[14]

## 6. Frameworks of reviews helpfulness.

Many of frameworks in this field proposed in recent years. In this section, we will review some of these frameworks as we see in

The authors in [34] proposed a ‘Sentiment Analysis as a Service’ (SAaaS) framework. This framework extracted sentiments from social information services then examined and converted this data into helpful information. In addition, they suggested a dynamic assistance composition mechanism for sentiment analysis based on the social fact’s classification. As an evaluation, they conducted the experiment on real-world datasets and presented the performance of this framework.



Another framework proposed in [35], the researcher suggested framework called “Senpy”. This framework involved all the aspects of developing, publishing and using a sentiment analysis service. They classified the aspects into layers to help in transferring the modularity to its implementations.

The author in [36] proposed a framework for E-mail sentiment analysis. The framework used a mixture structure of algorithms combined with K-means clustering and support vector machine classifier (SVM). They evaluated the framework through difference between three labeling methods: SentiWordNet labeling, K-means labeling, and Polarity labeling. The framework used five classifiers: Support Vector Machine, Naïve Bayes, Logistic Regression, Decision Tree and OneR.

Another framework was proposed in [37] it is called implicit social trust and sentiment (ISTS). ISTS considered as recommender system based that defines user’s preferences by discovering the user’s online social networks. The idea of ISTS is based on personalize suggestions by mining short text post of friends of users and measuring the implicit trust between them based on intercommunication activity.

In addition, ISTS gathered the sentiment rating to reproduce the knowledge behind friends’ short posts by using sentiment techniques. Then it identified the effect degree of confidence level between friends and sentiment rating by using machine learning regression algorithms.

In [20] framework was proposed for analyzing the impact of helpful reviews on product sales. Different characteristics of customer reviews and their impact on product sales are hypothesized in proposed framework. They examined three perspectives include impact of Informative of helpful reviews, impact of message persuasiveness of helpful reviews and impact of valence and series of helpful reviews.

In [28] the researcher implemented a multilingual framework to forecast review helpfulness. The framework abstracted review helpfulness as a combined of reviewer, review and product attributes. In detailed the use the following variables: number of helpful votes, reviewer Engagement, number of reviews, user compliments, user friend number, Product review valence, product review volume, product number check-in, review sentiment score, review length.

## **7. Evaluation method for review helpfulness prediction.**

Many of evaluation methods are used for assessing the validity of models as we see in

.In detail for example the authors in [38] applied ten-fold cross-validation to all trial estimations. In addition, they considered the metrics of accuracy, sensitivity, specificity, precision, recall, confusion matrix, and F-measure to estimate the model performance.

In addition the model in [22] was evaluated by measure whether the model accomplishes the major objective: finding the most-helpful reviews. Therefore, they observed the accuracy of their proposed model.

Then, they progressed in different stage of evaluation process. First: calculated the review helpfulness of the suggested model. Second, categorized the reviews depending on their estimated helpfulness. Third, decided how strictly the estimated category differs from the real category based on the users' helpfulness votes. Fourth, computed the mean absolute error MAE (mean difference between the estimated and the real category) to assess the proposed model.

On the other hand, the authors in [16] observed the transformation between their model and other models based on a view of the market, they selected a Tobit regression model in [39] as a benchmark and produced snapshots of the reviews every day for 40 day. Then calculated the Tobit model on the records of each view and displayed the p-values of the coefficients of the static attributes.

While the researcher in [11] validated the performance by composed the ANN MLP in three layers: an input layer, one disappeared layer and an output layer. To put suitable number of neurons in the disappeared layer they used Cross-validation method and used the logistic function as start and a linear activation function for the output layer.

In [10] the evaluation done by compare the model with two illustrative one stage models on review helpfulness in the literature. They used same methods in identifying the independent variables except some changes in the valence specification.

Moreover, the authors in [23] utilized the following error-based measures to validate regression performance: root relative squared error (RRSE), relative absolute error (RAE), root mean squared error (RMSE), and mean absolute error (MAE).

Finally, in [28] the researcher assessed the effectiveness of proposed model by focused on its capability to handled multilingual review text and produced related numeric variables. They compared a standard model versus their proposed model specification. They used RapidMiner19 to calculate and estimate the performance

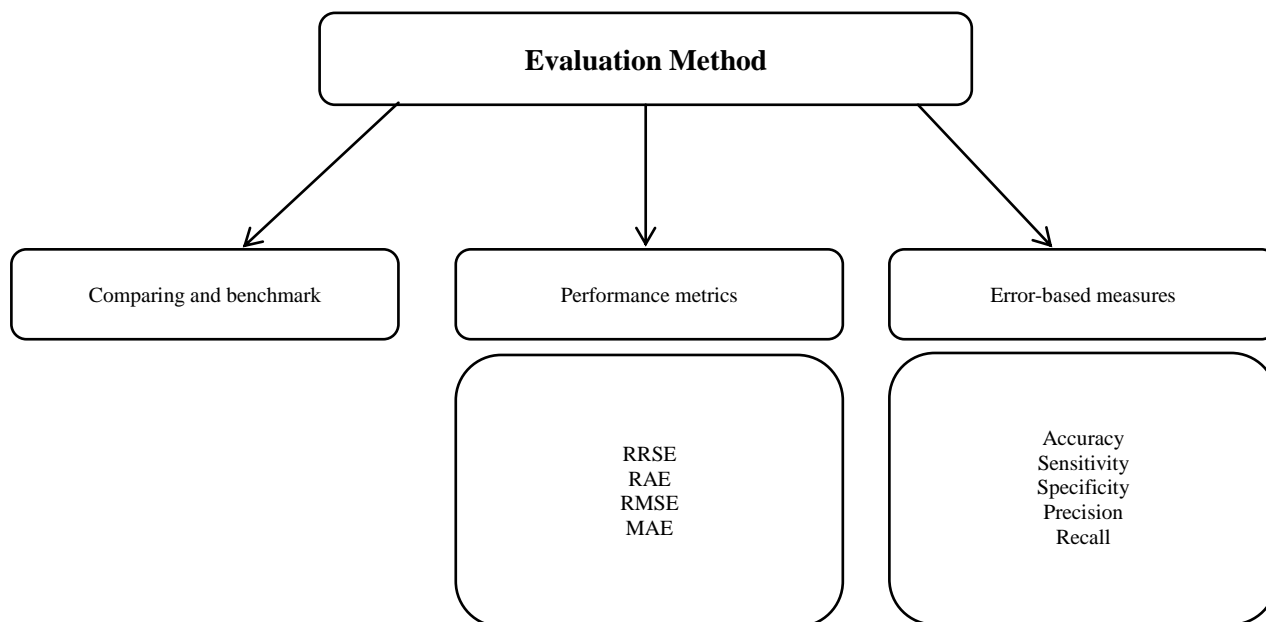
of the standard and proposed models. They split the dataset into two equal samples training dataset to adjust the model and the examining dataset to calculate the accuracy of the prediction.

**Table 4** highlight the main evaluation method used in the literature.

**Table 4: Evaluation method for review helpfulness prediction.**

Evaluation method	Evaluation type	References
Performance metrics	Accuracy	[38]
	Sensitivity	
	Specificity	
	Precision	
	Recall	
	F-measure	
	Confusion matrix	
Error-based measures	Root relative squared error (RRSE)	[23] [22]
	Relative absolute error (RAE)	
	Root mean squared error (RMSE)	
	Mean absolute error (MAE).	
Comparing and benchmark		[16] [22] [28]

In addition, the following Figure 2 clarify the most used evaluation method in the review helpfulness experiments. As we mention before, we notice the majority of the evaluation methods are between performance metrics and comparing or benchmark to other experiments.



**Figure 2: Most evaluation method used in review helpfulness prediction.**

### 8. Discussion.

According to previous discussion of some attributes of review helpfulness. We observed that the most critical attributes effect the review helpfulness is Review Valence (Review Sentiment). It was performed high coefficients and accuracy in many papers because it has direct relationships to the review helpfulness. In

addition, in reviewer's attributes side, we found the reviewers behavior attribute have better relationship to predict the reviews helpfulness.

On the other side, the most datasets used in previous researches are gathered from Amazon.com website. The reason of that is because of huge numbers of reviews and reviewers in comparison with other websites. Besides that, amazon provides good enough reviewers attributes that can use for different prediction experiments. This makes Amazon reviews considered as a powerful source of information for academic researchers in the fields of prediction and classification.

For the techniques used in review helpfulness prediction, we noticed that most of the previous researches focus on regression method. For example, many of experiments in the literature used regression such as Tobit regression, linear regression and support vector regression (SVR). The main reason for its commonly used is its ability to understand how the value of the dependent variable varies when any one of the independent variables is changed, although the other independent variables are stable [40]. The other technique used mostly in the literature is correlation coefficients which is a numerical measure of some type of correlation between two variables [41]. The main advantage of correlation coefficients is simple but accurate for calculate correlation coefficients between variable and a group of independent variables [42].

In addition, the mostly evaluation methods used in the literature are performance measurements include accuracy, sensitivity, specificity, precision, recall and confusion matrix. These methods are commonly used due to its ability to classify the accurate classification of labels within different classes and focus on one class such as positive [43].

## 9. Conclusion

The review helpfulness can be affected by many attributes. Some of them are related to (1) review text itself and, (2) other are related to reviewers. This paper discussed the main attributes that have major effect on review helpfulness. We chose some important attributes in both perspectives and revised their impact on review helpfulness. Then other dimensions such as datasets, techniques, frameworks and evaluation methods are listed and discussed from many papers. In addition, we highlighted the most important findings from all perspective. As future work, we recommend addressing some combination of many attributes from both sides (review and reviewers). Based on the previous literature, most of the experiment focus on one attribute in same time. However, making some combination of two or more attributes can boost the review helpfulness prediction. This point will be addressed in future thesis of the main author.

## References

1. Moore, S.G., *Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions*. Journal of Consumer Research, 2015. **42**(1): p. 30-44.
2. Baek, H., J. Ahn, and Y. Choi, *Helpfulness of online consumer reviews: Readers' objectives and review cues*. International Journal of Electronic Commerce, 2012. **17**(2): p. 99-126.
3. Cao, Q., W. Duan, and Q. Gan, *Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach*. Decision Support Systems, 2011. **50**(2): p. 511-521.
4. Salehan, M. and D.J. Kim, *Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics*. Decision Support Systems, 2016. **81**: p. 30-40.
5. Filieri, R., *What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM*. Journal of Business Research, 2015. **68**(6): p. 1261-1270.
6. Yin, D., S. Bond, and H. Zhang, *Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews*. 2013.
7. Ghose, A. and P.G. Ipeirotis, *Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics*. IEEE Transactions on Knowledge and Data Engineering, 2011. **23**(10): p. 1498-1512.
8. Kim, S.-M., et al. *Automatically assessing review helpfulness*. in *Proceedings of the 2006 Conference on empirical methods in natural language processing*. 2006. Association for Computational Linguistics.
9. Li, M., et al., *Helpfulness of online product reviews as seen by consumers: Source and content features*. International Journal of Electronic Commerce, 2013. **17**(4): p. 101-136.
10. Wu, J., *Review popularity and review helpfulness: A model for user review effectiveness*. Decision Support Systems, 2017. **97**: p. 92-103.

11. Barbosa, J.L., R.S. Moura, and R.L.d.S. Santos. *Predicting Portuguese Steam Review Helpfulness Using Artificial Neural Networks*. in *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web*. 2016. ACM.
12. Zhang, Y. and D. Zhang. *Automatically predicting the helpfulness of online reviews*. in *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on*. 2014. IEEE.
13. Ahmed, E., et al. *Challenges, comparative analysis and a proposed methodology to predict sentiment from movie reviews using machine learning*. in *Big Data Analytics and Computational Intelligence (ICBDAC), 2017 International Conference on*. 2017. IEEE.
14. Singh, W. *Sentiment analysis of online mobile reviews*. in *Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on*. 2017. IEEE.
15. Gao, B., N. Hu, and I. Bose, *Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews*. *Decision Support Systems*, 2017. **95**: p. 1-11.
16. Lu, S., J. Wu, and S.-L.A. Tseng, *How Online Reviews Become Helpful: A Dynamic Perspective*. *Journal of Interactive Marketing*, 2018. **44**: p. 17-28.
17. Hong, H., et al., *Understanding the determinants of online review helpfulness: A meta-analytic investigation*. *Decision Support Systems*, 2017. **102**: p. 1-11.
18. Wang, X., L.R. Tang, and E. Kim, *More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness?* *International Journal of Hospitality Management*, 2018.
19. Pennebaker, J.W., et al., *The development and psychometric properties of LIWC2015*. 2015.
20. Kaushik, K., et al., *Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon*. in. *Journal of Retailing and Consumer Services*, 2018. **45**: p. 21-32.
21. Chen, Y., et al., *Analysis of review helpfulness based on consumer perspective*. *Tsinghua Science and Technology*, 2015. **20**(3): p. 293-305.
22. Siering, M., J. Muntermann, and B. Rajagopalan, *Explaining and predicting online review helpfulness: The role of content and reviewer-related signals*. *Decision Support Systems*, 2018. **108**: p. 1-12.
23. Ngo-Ye, T.L. and A.P. Sinha, *The influence of reviewer engagement characteristics on online review helpfulness: A text regression model*. *Decision Support Systems*, 2014. **61**: p. 47-58.
24. Goswami, K., Y. Park, and C. Song, *Impact of reviewer social interaction on online consumer review fraud detection*. *Journal of Big Data*, 2017. **4**(1): p. 15.
25. Karimi, S. and F. Wang, *Online review helpfulness: Impact of reviewer profile image*. *Decision Support Systems*, 2017. **96**: p. 39-48.
26. Abd-Elhamid, L., D. Elzanfaly, and A.S. Eldin. *Feature-based sentiment analysis in online Arabic reviews*. in *Computer Engineering & Systems (ICCES), 2016 11th International Conference on*. 2016. IEEE.
27. Mandal, S. and S. Gupta. *A Lexicon-based text classification model to analyse and predict sentiments from online reviews*. in *Computer, Electrical & Communication Engineering (ICCECE), 2016 International Conference on*. 2016. IEEE.
28. Zhang, Y. and Z. Lin, *Predicting the helpfulness of online product reviews: A multilingual approach*. *Electronic Commerce Research and Applications*, 2018. **27**: p. 1-10.
29. McAuley, J. and J. Leskovec. *Hidden factors and hidden topics: understanding rating dimensions with review text*. in *Proceedings of the 7th ACM conference on Recommender systems*. 2013. ACM.
30. Field, A.P., *Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed-and random-effects methods*. *Psychological methods*, 2001. **6**(2): p. 161.
31. Eibe Frank, M.A.H., and Ian H. Witten, h. *Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016. **Fourth Edition**.
32. Arellano, M. and O. Bover, *Another look at the instrumental variable estimation of error-components models*. *Journal of econometrics*, 1995. **68**(1): p. 29-51.
33. Yin, D., S. Bond, and H. Zhang, *Dreading and ranting: The distinct effects of anxiety and anger in online seller reviews*. 2011.
34. Ali, K., et al. *Sentiment analysis as a service: a social media based sentiment analysis framework*. in *2017 IEEE International Conference on Web Services (ICWS)*. 2017. IEEE.
35. Sánchez-Rada, J.F., et al. *Senpy: A pragmatic linked sentiment analysis framework*. in *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*. 2016. IEEE.

36. Liu, S. and I. Lee. *A hybrid sentiment analysis framework for large email data*. in *Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on*. 2015. IEEE.
37. Alahmadi, D.H. and X.-J. Zeng, *ISTS: Implicit social trust and sentiment based approach to recommender systems*. *Expert Systems with Applications*, 2015. **42**(22): p. 8840-8849.
38. Lee, P.-J., Y.-H. Hu, and K.-T. Lu, *Assessing the helpfulness of online hotel reviews: A classification-based approach*. *Telematics and Informatics*, 2018. **35**(2): p. 436-445.
39. Mudambi, S.M. and D. Schuff, *Research note: What makes a helpful online review? A study of customer reviews on Amazon.com*. *MIS quarterly*, 2010: p. 185-200.
40. Bolin, J.H., Hayes, Andrew F.(2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: The Guilford Press. *Journal of Educational Measurement*, 2014. **51**(3): p. 335-337.
41. McGraw, K.O. and S.P. Wong, *Forming inferences about some intraclass correlation coefficients*. *Psychological methods*, 1996. **1**(1): p. 30.
42. Meng, X.-L., R. Rosenthal, and D.B. Rubin, *Comparing correlated correlation coefficients*. *Psychological bulletin*, 1992. **111**(1): p. 172.
43. Sokolova, M., N. Japkowicz, and S. Szpakowicz. *Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation*. in *Australasian joint conference on artificial intelligence*. 2006. Springer.