



Swansea University  
Prifysgol Abertawe



## Cronfa - Swansea University Open Access Repository

---

This is an author produced version of a paper published in :  
*Journal of Business Research*

Cronfa URL for this paper:  
<http://cronfa.swan.ac.uk/Record/cronfa29584>

---

### **Paper:**

Singh, J., Irani, S., Rana, N., Dwivedi, Y., Saumya, S. & Kumar Roy, P. (2016). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*  
<http://dx.doi.org/10.1016/j.jbusres.2016.08.008>

---

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.  
<http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/>

## **Predicting the “helpfulness” of online consumer reviews**

Jyoti Prakash Singh, National Institute of Technology Patna, India, Email: [jps@nitp.ac.in](mailto:jps@nitp.ac.in)

Seda Irani, School of Management, Swansea University Bay Campus, Swansea SA1 8EN, UK, Email: [sedatuna@hotmail.com](mailto:sedatuna@hotmail.com)

Nripendra P. Rana, School of Management, Swansea University Bay Campus, Swansea SA1 8EN, UK, Email: [nrananp@gmail.com](mailto:nrananp@gmail.com)

Yogesh K. Dwivedi\*, School of Management, Swansea University Bay Campus, Swansea SA1 8EN, UK, Email: [ykdwivedi@gmail.com](mailto:ykdwivedi@gmail.com)

Sunil Saumya, National Institute of Technology Patna, India, Email: [sunils.cse15@nitp.ac.in](mailto:sunils.cse15@nitp.ac.in)

Pradeep Kumar Roy, National Institute of Technology Patna, India, Email: [pradeep.cse15@nitp.ac.in](mailto:pradeep.cse15@nitp.ac.in)

Submission: 31<sup>st</sup> Jan 2016

First Revision: 20<sup>th</sup> May 2016

Acceptance: ...

Send correspondence to: Yogesh K. Dwivedi, School of Management, Swansea University Bay Campus, Swansea SA1 8EN, UK, Email: [ykdwivedi@gmail.com](mailto:ykdwivedi@gmail.com)

# Predicting the “helpfulness” of online consumer reviews

## Abstract

Online shopping is increasingly becoming people’s first choice when shopping, as it is very convenient to choose products based on their reviews. Even for moderately popular products, there are thousands of reviews constantly being posted on e-commerce sites. Such a large volume of data constantly being generated can be considered as a big data challenge for both online businesses and consumers. That makes it difficult for buyers to go through all the reviews to make purchase decisions. In this research, we have developed models based on machine learning that can predict the helpfulness of the consumer reviews using several textual features such as polarity, subjectivity, entropy, reading ease etc. The model will automatically assign helpfulness values to an initial review as soon as it is posted on the website so that the review gets a fair chance of being viewed by other buyers. The results of this study will help buyers to write better reviews and thereby assist other buyers in making their purchase decisions, as well as help businesses to improve their websites.

**Keywords:** Online User Reviews, Helpfulness, Product Features, Text Mining, Product Ranking

## 1. Introduction

Online user reviews have become today's word of mouth for the current generation of customers and business managers. Hence, understanding the role of online user reviews in e-commerce has attracted the attention of both academics and practitioners around the world (Duan, Gu, & Whinston, 2008a; Forman, Ghose, & Wiesenfeld, 2008; Li & Hitt, 2008). Online user reviews influence both product sales via consumer decision-making and quality improvement via business firms (Duan, Gu, & Whinston, 2008b). With the rapid penetration of the Internet into society and e-commerce business, the amount of user reviews is increasing rapidly. Such a large volume of data constantly being generated can be considered as a big data challenge (Chen, Chen, Du, Li, Lu, Zhao, & Zhou, 2013) for both online businesses and consumers.

Online reviews in the form of unstructured big data have both negative and positive impacts on consumers. First of all, the consumers are getting the real experience of their peers about a product, which helps them make intelligent decisions about the product or service. But at the same time, the large amount of reviews can cause information overload. In some cases, it is not possible for any customer to go through all the reviews and then make decisions. For example, an average-ranked book on Amazon.com can have more than several hundred reviews, whereas for a popular product such as the latest mobile phone, the number of reviews can be in the thousands. In such situations, it is virtually impossible for consumers to read all the reviews before making purchase decisions, especially for products that have been reviewed by hundreds and sometimes thousands of customers with their inconsistent opinions. Chen et al. (2013) classify such a large volume of unstructured data (i.e., big data) in the form of user generated content, which clearly poses a big data management challenge. It would be more useful for customers if they had a higher level of visibility of helpful user reviews that reflect the overview of the product or services. That would encourage websites

to evaluate the helpfulness of reviews written by other users. This is traditionally done by asking a simple question such as “Was this review helpful to you?” and putting “thumbs up” and “thumbs down” buttons.

The usefulness of reviews is generally assessed and their rank assigned by websites based on the helpfulness voting. For example, by default, user reviews are sorted by their helpfulness on Amazon.com. This is very useful to consumers as they can see the most helpful reviews on top. This also makes the website more user-friendly and hence attracts more consumers. Reviews that are perceived as helpful to customers bring considerable benefits to companies, including increased sales (Chevalier & Mayzlin, 2006; Clemons, Gao, & Hitt, 2006). It is estimated that this simple question “Was this review helpful to you?” brings in about \$2.7 billion additional revenue to Amazon.com (Spool, 2009). However, the helpfulness voting is not a silver bullet and does not solve all problems. The reasons for this are (i) very few user reviews receive helpfulness votes, and without helpfulness votes, the helpfulness voting mechanism does not work effectively; and (ii) recent reviews have yet to get votes, and hence their helpfulness cannot be decided. Given that reviews are posted so rapidly, the useful reviews are likely to get buried beneath less useful reviews before attracting helpfulness votes.

Since most helpful reviews get higher exposure to consumers, they normally become more helpful whereas less helpful reviews become less attractive to consumers due to less exposure. As a result, the reviews with fewer helpfulness votes are ignored by customers whereas reviews with more helpfulness votes get more visibility and readership. The result of this is that consumer decision-making is mostly influenced by the helpfulness votes and is skewed without considering when the review was posted and what the context was.

Although online reviews have helped consumers in deciding the pros and cons of different products, which ultimately helps in deciding the best product for an individual's needs, they

introduce a challenge for consumers to analyze this huge amount of data because of its volume, variety, and velocity. Review data is getting big day by day, at a very fast pace. Some users have started putting up pictures/images of the product to make their reviews more attractive and user-friendly. Hence, the review dataset may be seen as a big data analytics problem. It is interesting for businesses to dig into those review data to get insight about their products. Chen et al. (2013, p.160) suggest that “machine learning is possibly a feasible way to improve traditional data reduction techniques to process or even pre-process big data”. However, such an approach (i.e., machine learning) is yet to be tried and tested for enhancing the value of online user reviews.

Considering the discussion presented above about using a machine learning approach for big data analysis, this research investigates the helpfulness of online consumer reviews. We propose a system where the website itself would be able to perform the initial evaluation of the review using the model put forward by this research. That would help in prioritizing the better reviews in an appropriate order so that they can be viewed by other users. This will mitigate the Matthew effect, which implies that the top reviews gain more helpfulness votes as they are more visible and the lower reviews get fewer helpfulness votes as they are buried inside the review heap (Wan, 2015). A recent study by *brightlocal.com* (BrightLocal, 2016) suggests that 87% of buyers read 10 or less than 10 reviews before trusting a business. Hence, if a review is really helpful, but it is not put in the top 10 list, then it will lose its purpose. The proposed approach ensures that this helpful review is ranked appropriately in the review.

We chose the Indian context for this study because e-commerce businesses have just started flourishing here. People have started buying online and writing reviews for the related products. The reviews on Indian e-commerce sites are very different from the reviews from other parts of the world where e-commerce is very popular. Reviews written by Indian buyers are mainly in English, but they contain some Hindi text (written in English script only) as

well. Some of the most widely used Hindi words, such as *bahut achha*, *bakbas*, and *pesa wasool*, are found in a number of reviews.

The rest of the article is structured as follows: Section 2 reviews the related literature and is followed by the methods of data collection and analysis in Section 3. The results are presented in Section 4. Next, Sections 5 and 6 report the discussion and conclusions respectively.

## **2. Literature Review**

Various research studies have been done on helpfulness of reviews. Some researchers have used regression techniques to show the most helpful reviews while others have used neural networks. Ghose and Ipeirotis (2006) propose two ranking mechanisms for ranking product reviews: a consumer-oriented ranking mechanism that ranks the reviews according to their expected helpfulness, and a manufacturer-oriented ranking mechanism that ranks them according to their expected effect on sales. They used econometric analysis with text mining to make their ranking work. They found the reviews that tend to include a mix of subjective and objective elements are considered more informative (or helpful) by the users.

Liu, Cao, Lin, Huang, & Zhou (2007) consider the product review helpfulness problem as a binary classification problem. They performed manual annotation to check review comments on many products using ‘favorable’ and ‘unfavorable’ as the classification targets but they did not use the original helpful feedback for their study. However, Liu, Huang, An, & Yu (2008) propose a model for predicting the helpfulness of reviews using many features, such as length of reviews and the writing style of the reviewers. Out of these, the three most important factors named and used for prediction are the reviewer’s expertise, the writing style of the reviewer, and the timeliness of the review. Radial basis functions are used to model expertise and writing style. The training data is taken from a tally present in the reviews itself, called a helpfulness vote.

Forman et al. (2008) suggest that in the context of an online community, the reviewer's disclosure of identity-descriptive information is used by consumers to supplement or replace product information when making purchase decisions and evaluating the helpfulness of online reviews. They found that the online community member's rate reviews containing identity-descriptive information more positively. Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee (2009) find a new correlation between proportion of helpful votes of reviews and deviation of the review ratings from the average ratings of products. They report that helpful votes are consistent with average ratings. Mudambi and Schuff (2010) undertake an analysis on reviews collected from Amazon.com to determine which properties of a review make it useful for the customers. Three hypotheses were formulated and verified. On analyzing the hypotheses, they found that the impact of review extremity was dependent on product type. Products were grouped into two categories: (i) search product and (ii) experience product. A search product is one that customers can easily acquire information about concerning its quality before interacting with the product directly and where it does not require much customer involvement' to evaluate the key quality attributes of the product, which are objective and easy to compare. An experience product is one that customers have to interact directly with to acquire information about its quality. With an experience product, the customer's involvement is required in order to evaluate the level of quality as key attributes are subjective or difficult to compare.

For experience products, extreme reviews were found to be less helpful as compared with moderate reviews. However, for search products, extreme reviews were more helpful than moderate ones. The review length also had an impact on helpfulness but was dependent on product type. For search products, review length had a greater positive impact as compared to experience products. So, it was concluded that helpfulness depended on star rating and review length but was also dependent on product type.

Ghose and Ipeirotis (2011) analyze many characteristics of review texts, such as spelling errors, readability, and subjectivity, and examined their impact on sales. Linguistic correctness was found to be a vital factor affecting sales. There is a feeling that reviews of medium length with fewer spelling errors are more helpful to naive buyers as compared to reviews that are very short or very long and have spelling errors.

To analyze the impact of various characteristics of online user reviews, Cao, Duan, and Gan (2011) use text mining on the helpfulness as indicated by the number of votes a reviewer receives. They found that helpfulness is more affected by semantic features as compared to other features of reviews. They also found that reviews expressing extreme opinions are more impactful than reviews with neutral or mixed opinions.

Korfiatis, García-Bariocanal, and Sánchez-Alonso (2012) explore the interplay between online review helpfulness, rating score, and the qualitative characteristics of the review text as measured by readability tests. They constructed a theoretical model based on three elements: conformity, understandability, and expressiveness. They investigated the directional relationship between the qualitative characteristics of the review text, review helpfulness, and the impact of review helpfulness on the review score. To validate their model, they used a dataset containing 37,221 reviews collected from Amazon UK. They found that review readability had a greater effect on the helpfulness ratio of a review than its length.

The sentiments of the review have a direct impact on sales as stated by several researchers (Li & Wu, 2010; Liu, Yu, An, & Huang, 2013; Schumaker, Zhang, Huang, & Chen, 2012; Siering & Muntermann, 2013) revealed a very unique property of reviews, indicating that reviews with information related to the quality of the product received more helpfulness votes. Sentiments are examined vigorously in the review analysis context, and the main aim of sentiment analysis or opinion mining is to extract the sentiment of the user regarding

products or features of a product. Liu et al. (2013) propose a new sentiment analysis model, in which a review is generated under the influence of a number of hidden sentiment factors. The parameters of the model, named S-PLSA (Sentiment Probabilistic Latent Semantic Analysis), gets updated as new review data becomes available.

Sparks, Perkins, & Buckley (2013) conduct an analysis on online travel reviews, unlike other researchers, who performed research on online shopping reviews. They found that travelling decisions, such as where to go, what to eat, where to stay, and places to visit, were more influenced by consumer-generated reviews than suggestions already provided by the travel and tourism agency.

In most of the research works, only customers' reviews were considered for the determination of various parameters. But Wu, Wu, Sun, & Yang (2013) carry out an analysis on both seller and customer reviews. Before purchasing any item, customers go through various things, such as customer reviews, seller reviews, and price comparison with other marketplaces. The authors used all these parameters to determine the willingness to pay of customers using a conceptual model.

Li, Huang, Tan, & Wei (2013) analyze content-based and source-based review features that directly influence product review helpfulness. It was also found that customer-written reviews that were less abstract in content and highly comprehensible result in higher helpfulness. Wang, Zhu, & Li (2013) propose a technique called "SumView", a web-based review summarization system to automatically extract the most representative expression of customer opinion in the reviews on various product features. In which a crawler is used to obtain the product review from the website and when a product id is given then all the reviews and comments given by the customer are downloaded.

Hu, Koh, and Reddy (2014) develop model containing multiple equations in order to determine the interrelationship between rating, sentiments and sales. Firstly, they undertook

sentiment analysis on the reviews, followed by an examination of the impact of sentiment on sales. Moderate sentiment was found to have stronger impact on sales. Moderately negative and positive sentiments had more impact compared to strongly positive and negative sentiments. This peculiar result could be due to the fact that reviews that were more positive or negative may have been manipulated. The rating was not found to have any direct impact on sales. Another important observation was that the most helpful and most recent reviews had a very strong impact on sales as these are the most accessible ones.

Lee and Shin (2014) investigate whether the quality of reviews affects the evaluations of the reviewers and the e-commerce website itself. They conducted pilot tests prior to the main experiment. The participants were asked questions such as (a) how frequently they use online shopping malls, and (b) if they had ever used the target product. They investigated (a) how the reader's acceptance depends on the quality of online product reviews and (b) when such effects are more or less likely to occur. Their findings indicated that participants' intention to purchase the product increases with positive high-quality reviews as opposed to low-quality ones.

Wan (2015) analyze a dataset of bestselling products of Amazon.com and emphasized the Matthew effect (Merton, 1968) and the ratchet effect (Freixas, Guesnerie, & Tirole, 1985). Reviews are no doubt helpful to the user who wants to know more about any product. The helpfulness of any review has become a major area of research, influenced by Amazon starting a new practice called the 'helpfulness vote'. As per 'helpfulness vote', Amazon ranks every review, and a few top reviews are shown on the product page. The most helpful review is also selected. The Matthew effect states that the top reviews gain even more helpfulness votes and the lower-ranked reviews get fewer helpfulness votes. The Matthew effect is analogous to the social Matthew effect of the 'rich becoming richer and poor becoming poorer'. The outcome, that the top review receives more attention and remains on the top for

a long time, is called the ratchet effect. This is a kind of biasing through which top reviews remain on the top. This is an inherent limitation of the reviews, which is contrary to their original purpose. Wan (2015) confirms the presence of the Matthew effect by analyzing the reviews of bestselling products of Amazon.com.

Krishnamoorthy (2015) examines the factors influencing the helpfulness of online reviews and built a predictive model. His proposed predictive model extracts linguistic features such as adjectives, state verbs, and action verb features and accumulates them to make linguistic feature (LF) value. He also used review metadata (review extremity and review age), subjectivity (positive and negative opinion words), and readability-related (Automated Readability Index, SMOG, Flesch–Kincaid Grade Level, Gunning Fog Index, and Coleman–Liau Index) features in their model for helpfulness prediction.

Huang, Chen, Yen, and Tran (2015) examine message length together with aspects of review patterns and reviewer characteristics for their joint effects on review helpfulness. They found that the message length in terms of word count has a threshold in its effects on review helpfulness. Beyond this threshold, its effect diminishes significantly or becomes near non-existent. Allahbakhsh, Ignjatovic, Motahari-Nezhad, and Benatallah (2015) propose a set of algorithms for robust computation of product rating scores and reviewer trust ranks. They harvested user feedback from social rating systems. Social rating systems collect and aggregate opinions (experience of using a service, purchasing a product, or hiring a person that is shared with other community members, in order to help them judge an item or a person that they have no direct experience with) to build a rating score or level of trustworthiness for items and people. This paper introduced a compatible framework, which consists of three main components that are responsible for (1) calculating a robust rating score for products, (2) behavior analysis of reviewers, and (3) trust computation of reviewers. In the presence of

unfair reviews, they proposed a novel algorithm for calculating robust rating scores for products.

Chua and Banerjee (2016) find a relation between helpfulness and review sentiment, helpfulness and product type, and helpfulness and information quality. Review sentiment was classified in three categories: favorable, unfavorable, and mixed. The products were categorized as search products and experience products. The information quality has three major dimensions: comprehensibility, specificity, and reliability. Comprehensibility refers to the understandability of reviews, specificity refers to the adequacy of information given in reviews, and reliability refers to the dependency of consumers on reviews (Chua & Banerjee, 2016). By analyzing various data from Amazon.com, they concluded that helpfulness varied across review sentiment and was independent of product type. However, information quality and helpfulness varied as a function of both product type and sentiment of the review. Qazi, Syed, Raj, Cambria, Tahir, and Alghazzawi (2016) explain why some reviews are more helpful compared to others. As the helpfulness of online reviews helps the online web user to select the best product, they read several reviews of that product and finally conclude whether the review was helpful or not.

The analysis of reviews includes various aspects such as sentiment analysis and helpfulness calculation. For performing such an analysis, several techniques are used, including classification techniques, Naïve Bayes theorem, support vector machines, natural language processing, and regression techniques. Some researchers claim that natural language processing is more efficient whereas others support regression techniques. We are using a new technique called as ensembles learning technique. Moreover, in prior research, most of the analyses were performed on a very small dataset. Only a few studies such as Hu, Bose, Koh, and Liu (2012) have taken large data, but the data lacks variety. A summary of the relevant research on review helpfulness is presented in Table 1.

### Table 1 here

Researchers have observed the Matthew and the ratchet effect, which hinders the accurate calculation of helpfulness. By calculating helpfulness on the basis of various parameters using machine learning techniques, we have mitigated these effects. We have collected around six hundred thousand reviews, which contain data from baby products, electronic products, and books.

## 3. Research Methodology

### 3.1. Data collection

Review data were collected from Amazon.in for three categories of products, namely books, baby products, and electronic products, with raw data size of 3 GB, 437 MB, and 530 MB respectively. We used data scraping to collect the data. Data scraping is a technique by which a computer program extracts data from human-readable output coming from a website. Data scraped from Amazon.in were then filtered and pre-processed to collect the following fields, which were Amazon Standard Identification Number (ASIN): the alphanumeric product ID given by Amazon to its product; **Reviewer ID**: the alphanumeric identification given by Amazon to its reviewers; **Reviewer Name**: the screen name of the reviewer on Amazon.in given by the reviewer only; **Title**: the title of the review given by the reviewer; **Review Time**: the date and time the review was written; **Helpfulness**: the total helpfulness votes received by the review; and **Review Text**: the review text written by the reviewer. A typical review on Amazon.in is given here in the figure below. The **ASIN** and **Reviewer ID** are not shown here, but they can be extracted from the scraped data.

Some of the reviews were discarded as due to the very small amount of review text they hardly gained any helpfulness votes. Such reviews were not very informative either. After discarding those data, we worked with 622,494 reviews, out of which 171,082 were for

books, 232,936 for baby products, and 218,477 for electronic products respectively. The custom-developed program in Python was then used to extract different variables we were going to use for helpfulness prediction. The following information was extracted from the review text: *Polarity, Subjectivity, Noun, Adjective, Verb, Flesch reading, Dale\_Chall\_RE, Difficult\_Words, Length, Set\_Length, Stop\_Count, Wrong\_Words, One\_letter\_Words, Two\_Letter\_Words, Longer\_letter\_Words, Lex\_Diversity, Entropy, Rating, and Helpful\_Ratio.*

**Figure 1 here**

### 3.2. Variables

The dependent variable for our model was *helpfulness ratio*, which is the percentage of people who vote for helpfulness against the total number of votes. Amazon.in does not show the total number of reviewers who voted for a product; it only shows the helpful votes for the product as shown in Figure 1. We took a slightly different approach to find the helpfulness ratio. Amazon.in does maintain a reviewer ranking list where it displays the *total reviews, helpful votes, and percent helpful* of a reviewer as shown in Figure 2. The *percent helpful* of a reviewer is the helpful votes divided by the total number of votes attracted by all the reviews written by that reviewer. So, we approximated the *helpfulness ratio* (the dependent variable) using the parameter *percent helpful* of a reviewer. For each review, we found the reviewer's name and then found her/his *percent helpful* to approximate the *helpfulness ratio* of the review.

The explanatory variables were *Polarity, Subjectivity, Noun, Adjective, Verb, Flesch\_Reading\_Ease, Dale\_Chall\_Readability, Difficult\_Words, Length, Set\_length, Stop\_count, Wrong\_words, One\_letter\_words, Two\_letter\_words, Longer\_letter\_words, Lex\_diversity, Entropy, and Rating.*

**Figure 2 here**

The *Polarity* of a text represents whether the expressed opinion in that text is negative, positive, or neutral (Wilson, Wiebe, & Hoffmann, 2009). The *Polarity* of the review text is measured by finding the total positive score minus total negative score of the review text. To calculate the positive and negative scores of a text, SentiWordNet (Esuli & Sebastiani, 2006) database is used. We developed a program in Python to find the *polarity* of the review text by consulting SentiWordNet. The textual information of reviews is generally categorized into two parts: (i) facts and (ii) opinions. Facts are objective expressions whereas opinions are subjective expressions (Liu, 2010). The *Subjectivity* is calculated by finding how many sentences of the review are expressing opinion and dividing this by the total number of sentences in the review (Liu, 2010). The *Nouns, Adjectives and Verbs* are calculated directly from the review by counting the respective numbers in it. The review text is tokenized (broken into words, phrases, symbols, or other meaningful elements called tokens) using Natural Language Toolkit (NLTK) (Bird, 2006). The readability indices such as *Flesch\_Reading\_Ease* (Kincaid, Fishburne, Rogers, & Chissom, 1975) and *Dale\_Chall Readability* (Dale & Chall, 1948; 1949) are calculated using the respective formulae as shown below.

The **Flesch Reading Ease** score is one of the best-known and most popular readability indicators. The formula for the Flesch Reading Ease score is given as:

$$\text{FRE} = 206.84 - (1.02 * \text{ASL}) - (84.6 * \text{ASW}) \quad (1)$$

Where,

FRE : Flesch Reading Ease readability score

ASL : Average sentence length in words or average number of words in a sentence (number of words divided by the number of sentences)

ASW : Average syllables per word (the number of syllables divided by the number of words)

**The Dale\_Chall formula:**

Unlike other formulae, which use ‘word-length’ to assess word difficulty, Dale-Chall uses a count of ‘Difficult words’. This makes the Dale-Chall formula unique (Chall & Dale, 1995).

The Dale-Chall formula is given as:

$$\text{Raw score} = 0.16*(\text{PDW}) + 0.05*(\text{ASL}) \quad (2)$$

Where,

PDW: Percentage of difficult words

ASL: Average sentence length in words

If PDW is greater than 5%, then:

$$\text{Adjusted score: Raw score} + 3.64;$$

Otherwise,

$$\text{Adjusted score: Raw score.}$$

The *Length* and *Set\_length* of review is measured by counting total words and unique words in the review text. The stop count and wrong words are calculated with respect to English text and are measured by consulting Enchant (Perkins, 2014) English dictionary. The *One\_letter\_word*, *Two\_letter\_word*, and *Longer\_letter\_words* are measured directly from the review text by counting the number of characters in each word. The *Stop\_words* are calculated for English language using NLTK (Bird, 2006). The *Lex\_diversity* is the ratio of *Set\_length* to the *Length* of the review. The *Entropy* measures in a certain sense how much information is produced on average for each word in the text. The *Entropy* (Shannon, 1951) of the text is calculated using the formula given by Equation 3

$$H(X|Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(y_j)}{p(x_i, y_j)} \quad (3)$$

Where  $p(x_i, y_j)$  represents probability. This quantity should be understood as the amount of randomness in the random variable  $X$  given the event  $Y$ . To the best of our knowledge, the *Entropy* measure has not been employed till now to find the helpfulness of reviews. From our

analysis, we found that *Entropy* happens to be one of the important parameters in deciding the helpfulness of reviews. A brief description of the variables is given in Table 2.

**Table 2 here**

### *3.3. Analysis Method*

We used an ensemble learning technique (gradient boosting algorithm) to analyze the data. Ensemble learning employs multiple base learners and combines their predictions. The fundamental principle of dynamic ensemble learning is to divide a large datastream into small data chunks. Then classifiers are trained on each data chunk independently. Finally, a heuristic rule is developed to organize these partial classified results into one super classified result.

This structure had many advantages. Firstly, each data chunk was relatively small so that the cost of training a classifier on it was not high. Secondly, we saved a well-trained classifier instead of the whole instances in the data chunk, which cost much less memory. Gradient boosting developed an ensemble of tree-based models by training each of the trees in the ensemble on different labels and then combining the trees. For a regression problem where the objective was to minimize MSE, each successive tree was trained on the errors left over by the collection of earlier trees.

The total dataset was divided into two parts (i) training set and (ii) testing set. 70% of the dataset was kept in the training set, and this dataset was used to train our system using a gradient boosting algorithm. The remaining 30% made up our testing dataset, which was used to test the performance of the system. Our objective in training was (i) to find the optimal number of trees in the ensemble, which minimizes the MSE, and (ii) find the ordering of variables influencing the helpfulness ratio. The findings are explained in the next section.

#### 4. Results

We found from our experiment that approximately 100 trees in the ensemble were giving the least MSE value for testing a dataset of books. The MSE for the training dataset of books was constantly decreasing with the increase in number of trees in the ensemble as shown in Figure 3. But when we analyzed the testing data of the books dataset, we found that the MSE starts increasing very slightly after a certain number of trees as shown in Figure 3. For the books dataset we found that approximately 100 trees are best for prediction as they give least MSE.

#### Figure 3 here

We found that rating of the reviews influenced the helpfulness ratio most in the books dataset. Since books are experience products, rating plays a major role. The next most influencing variables were found to be polarity and Dale\_Chall readability index as shown in Figure 4.

The high variable importance given for polarity indicates that customers expect some sentimental aspects to be highlighted in the reviews so that they can easily make decisions on whether to purchase the products or not. The high variable importance for polarity could lead to customers making both favorable and non-favorable decisions based on the positive or negative sentiments associated with a review. The relatively high variable importance for Dale\_Chall indicates that the reviewers have used relatively simple words so that the texts can be understood, and that it eventually helps the potential buyers in making decisions if the words used in reviews are selected from among 3000 words that could be effectively understood by any purchasers. Similarly, the higher entropy indicates that the meaning of each word used in the reviews written by customers is relatively suitable when describing the product.

Moreover, the findings also indicate that the variables such as adjectives, `difficult_words`, `set_length`, `wrong_words`, `lex_diversity`, `one_letter_words`, `verb`, `noun`, and `two_letter_words` attract the least attention of potential buyers in review comments.

**Figure 4 here**

The graph for MSE vs. number of trees in training and testing datasets for baby products is shown in Figure 5. It can be seen from the graph that the optimal number of trees for baby products dataset is also 100 as the MSE for the testing dataset is the least there.

**Figure 5 here**

As baby products are also experience products, potential buyers attribute more importance to the rating of the product given by experienced buyers, hence rating is a significant determiner for the helpfulness vote. Although rating comes in at third place in parameter ranking, its value is close to '1' on a scale of '0' to '1' as shown in Figure 6. The readability index of Dale\_Chall and polarity are found to be the most influencing parameters for helpfulness in this context.

**Figure 6 here**

FleschRE, subjectivity, and entropy are the other three influencing variables that have higher variable importance similar to books, which is the other type of experience product. Also similar to books, the variables such as adjective, `difficult_words`, `set_length`, `wrong_words`, and `lex_diversity` are given the least weightage in terms of making decisions to purchase baby products. The MSE vs. number of trees in ensemble for electronic products is shown in Figure 7. Similar to the other two testing set MSE, the best number of trees in the ensemble is 100 as the MSE for the testing dataset is least there.

**Figure 7 here**

Electronic products are an example of search goods, and we found that the rating does not affect the helpfulness as strongly compared to experience products such as baby products and

books. The Dale\_Chall readability index and polarity play the major role as determinants of helpfulness votes as the two highest important variables (see Figure 8). The findings clearly indicate that an easy-to-understand set of words and explicitly expressed positive, negative, or neutral opinions are equally important variables for search products such as electronic goods. Similar to books and baby products, electronic products also indicate FleschRE, subjectivity, and entropy among the top six variables of significant importance. Moreover, the variables of the least significant importance such as difficult\_words, adjective, wrong\_words, set\_length, verb, and one\_letter\_words are more or less similar to the other two products analyzed earlier.

### **Figure 8 here**

More precisely, the findings for all three products for Amazon.in clearly indicate that the top six variables for all cases are Dale\_Chall, polarity, FleschRE, subjectivity, entropy, and rating in a slightly different order. The result highlights that the readability of the text, which includes Dale\_Chall and FleschRE, is an important parameter for making a review helpful. The other important parameter is sentiment, which includes polarity, subjectivity, and ratings. People rate those reviews that contain more subjective statements and opinions highly, and this is in line with prior research (Cao et al., 2011; Ghose & Ipeirotis, 2011; Mudambi & Schuff, 2010). The other important parameter is entropy, which is normally not explored so much in the area of helpfulness analysis, and probably this is the first research where we have studied the effect of entropy on helpfulness votes.

Many researchers (e.g., Krishnamoorthy, 2015; Lee & Choeh, 2014) have mentioned one or more of the above parameters in their research work, but none of them have ranked these parameters. This is the first research where the parameters have been ranked and mapped to helpfulness. Here the ranking of parameters shows which parameters affect helpfulness more.

The parameter which is at the top is considered as the highest rank, and the parameter which is at the bottom has the lowest rank as shown in Figures 4, 6, and 8.

## 5. Discussion

Online reviews can be a powerful promotional tool for e-commerce websites, particularly when the huge amount of information available on the web has created information overload among online users (Cao et al., 2011; Hu et al., 2012; Lee & Choeh, 2014). Marketers and vendors have used this medium because it provides a cheap, effective, and impactful channel to reach their customers. Marketers take advantage of networks of experienced customers to influence the purchase behavior of potential buyers (Hu et al., 2012). The “helpfulness” feature of online user reviews makes it easier for consumers to cope with information overload and helps in their decision-making process (Cao et al., 2011; Krishnamoorthy, 2015).

The findings of this research indicate that approximately 100 trees in the ensemble give the least MSE value for testing datasets for all three different products (i.e., books, baby products, and electronic products) for Amazon as this number of trees is best for prediction. The MSE vs. number of trees for products datasets clearly indicates that even though the MSE decreases with a higher number of trees in the training dataset, it remains almost invariant irrespective of the increment in number of trees in the ensemble in the testing dataset. Therefore, it can be derived from the results that just 100 trees are enough to test the performance of the system using the testing dataset and are best for prediction as they give least MSE.

In order to extract useful information regarding the effect of various determinants on the helpfulness of online reviews, this research analyzed 19 variables encompassing the product data, review characteristics, and textual characteristics of the reviews. These variables are

considered important in how they affect the level of helpfulness (Lee & Choeh, 2014). The findings in terms of variable ranking for all three products analyzed clearly indicate that Dale\_Chall readability index, polarity, rating, FleschRE, subjectivity, and entropy are the most significant parameters to determine the helpfulness of online reviews. This outcome clearly indicates that ease of readability (shown by high Dale and FleschRE) with explicit positive or negative opinions and evaluations about the product expressed through meaningful and concise wording (shown by high polarity, entropy, and subjectivity) and supported with star ratings given for it make a review extremely helpful for potential buyers. Similar findings were also supported by the prior research (e.g. Cao et al., 2011; Krishnamoorthy, 2015).

The findings for this research also indicate that irrespective of the types of products, viz. search products and experience products (Huang, Lurie, & Mitra, 2009; Nelson, 1970), the helpfulness characteristics for different classes of products are almost the same (i.e., books and baby products as experience products in comparison to electronic items as search products). Huang and Yen (2013) claimed that helpfulness characteristics are different for these two classes of goods. While a helpful search product review is likely to contain more information on product aspects or attributes, a helpful experience product review is likely to have more descriptions of customer experiences (Krishnamoorthy, 2015). The high variable rankings of Dale\_Chall readability index, polarity, rating, FleschRE, subjectivity, and entropy for both search and experience products in our research clearly indicate that these variables are equally important, irrespective of their product types.

### *5.1. Implications for Business Research and Practice*

The results of this study have several implications for research in this field. This is the first study to predict the helpfulness of an online review using ensemble learning techniques. Ensemble learning is appropriate for estimating complex relationships among variables and

does not demand specific assumptions pertaining to the functional form or the distribution of the error terms. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms. Moreover, the difficulty posed by the great variation inherent in the review content and quality is alleviated. As consumer reviews have become a prominent source of product information, it is necessary for sites that publish online reviews to understand how these reviews are perceived by consumers. These results are in line with previous studies on the helpfulness of online reviews, which verified the importance of the review extremity variable (Cao et al., 2011; Mudambi & Schuff, 2010) or the level of informativeness based on textual characteristics (Ghose & Ipeirotsis, 2011). Thus, this study extends the earlier studies with its application of ensemble learning for verifying the effects of textual characteristics of review text on helpfulness.

Ensemble learning outperformed the conventional linear regression model analysis in predicting helpfulness, indicating that the proposed ensemble learning model has advantages when the model analyzes data with complex and nonlinear relationships between helpfulness and its determinants. The results of this study have implications for practitioners in that they offer several clues about site design for online reviews. The results of this study can be used to develop guidelines for creating more valuable online reviews. The design of user review systems to promote more helpfulness votes for online user reviews can be facilitated by enhancing the level of understanding of what drives helpfulness voting. This study explores the characteristics of online user reviews and how they influence the number of helpfulness votes.

## **6. Conclusion**

To find the helpfulness of a review, we have evaluated the different determinants of review helpfulness. The textual features such as readability, polarity, subjectivity, entropy,

and the average review rating of the product over time have been found to be the most important parameters for helpfulness. In addition, wrong words, stop words, length (number of words), and the number of one-letter words are other textual characteristics of reviews that are not so important parameters for helpfulness. Our system mitigates the Matthew effect and encourages the reviewers to write better reviews because the reviews get listed at a proper location on the review list as soon as they are posted on the website.

The other important finding of this research is that ensemble learning techniques are found to be better than linear regression techniques. They scale well to large-scale data and perform better. The MSE obtained when testing the data on various products using a gradient boosting technique, a type of ensemble learning, is much less than that obtained when using linear regression. For experience products, the star rating was found to be a more important parameter than for search products. But for both type of products, the readability, entropy, and sentiment parameters were found to have a similar effect.

### *6.1. Limitations and Future Research Directions*

There are some limitations of this research. First, one of the major limitations of this work is that the non-English words are counted as wrong words in this study, but a number of non-English words contain polarity information. These non-English words may be taken into account for polarity calculation in a future study. Second, this research has examined online reviews data relating to only one e-commerce website. Therefore, the findings of this research should be generalized to the other contexts consciously. Moreover, future research can also examine the “helpfulness” of online reviews for other websites as well, and the importance of variables considered in this research can then be compared to the different product types of any such e-commerce websites. Finally, this research uses more than six hundred thousand online reviews to explore and analyze nineteen variables and evaluate their

performance. Future research can consider even larger numbers of reviews and explore if any additional variables emerge from such big data analysis.

## References

- Allahbakhsh, M., Ignjatovic, A., Motahari-Nezhad, H. R., & Benatallah, B. (2015). Robust evaluation of products and reviewers in social rating systems. *World Wide Web*, *18*, 73-109.
- Amazon (2016). Coolpad Note 3 (Black, 16GB). Accessed from the web-link: [http://www.amazon.in/product-reviews/B0158IT7ES/ref=cm\\_cr\\_arp\\_d\\_viewopt\\_sr?ie=UTF8&showViewpoints=1&sortBy=helpful&filterByStar=four\\_star&pageNumber=1](http://www.amazon.in/product-reviews/B0158IT7ES/ref=cm_cr_arp_d_viewopt_sr?ie=UTF8&showViewpoints=1&sortBy=helpful&filterByStar=four_star&pageNumber=1), Accessed on 23<sup>rd</sup> May 2016.
- Bird, S. (2006). NLTK: The natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions* (pp. 69-72). Association for Computational Linguistics.
- BrightLocal (2016). Local Consumer Review Survey. Accessed from [www.brightlocal.com/learn/local-consumer-review-survey](http://www.brightlocal.com/learn/local-consumer-review-survey), on 16<sup>th</sup> May 2016.
- Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, *50*, 511-521.
- Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: A data management perspective. *Frontiers of Computer Science*, *7*, 157-164.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*, 345-354.
- Chua, A. Y., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, *54*, 547-554.

- Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23, 149-171.
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. *Educational Research Bulletin*, 37-54.
- Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26, 19-26.
- Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: A case study on amazon.com helpfulness votes. In Proceedings of the 18<sup>th</sup> International Conference on World Wide Web, 141-150.
- Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter?—An empirical investigation of panel data. *Decision Support Systems*, 45, 1007-1016.
- Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing*, 84, 233-242.
- Esuli, A., & Sebastiani, F. (2006). *Sentiwordnet: A publicly available lexical resource for opinion mining*. In Proceedings of LREC (Vol. 6, pp. 417-422).
- Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19, 291-313.
- Freixas, X., Guesnerie, R., & Tirole, J. (1985). Planning under incomplete information and the ratchet effect. *The Review of Economic Studies*, 52, 173-191.
- Ghose, A., & Ipeiritos, P. G. (2006). Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality. In Proceedings of the 16th Annual Workshop on Information Technology and Systems, 303-310.

- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23, 1498-1512.
- Hoang, L., Lee, J. T., Song, Y. I., & Rim, H. C. (2008). *A model for evaluating the quality of user-created documents*. In *Information Retrieval Technology*. Springer Berlin Heidelberg, 496-501.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177.
- Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52, 674-684.
- Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems*, 57, 42-53.
- Huang, A. H., & Yen, D. C. (2013). Predicting the helpfulness of online reviews—A replication. *International Journal of Human-Computer Interaction*, 29, 129-138.
- Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, 48, 17-27.
- Huang, P., Lurie, N. H., & Mitra, S. (2009). Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, 73, 55-69.
- Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. In *AAAI*, 22, 1331-1336.

- Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. *In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 423-430.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (No. RBR-8-75). *Naval Technical Training Command Millington TN Research Branch*.
- Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11, 205-217.
- Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications*, 42, 3751-3759.
- Lee, E. J., & Shin, S. Y. (2014). When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo. *Computers in Human Behavior*, 31, 356-366.
- Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, 41, 3041-3046.
- Li, M., Huang, L., Tan, C. H., & Wei, K. K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17, 101-136.
- Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48, 354-368.
- Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19, 456-474.

- Liu, B. (2010). *Sentiment Analysis and Subjectivity*. Handbook of Natural Language Processing, 2, 627-666.
- Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. *In EMNLP-CoNLL*, 334-342.
- Liu, Y., Huang, X., An, A., & Yu, X. (2008). Modelling and predicting the helpfulness of online reviews. *Eighth IEEE International Conference on Data Mining*, 443-452.
- Liu, Y., Yu, X., An, A., & Huang, X. (2013). Riding the tide of sentiment change: sentiment analysis with evolving online reviews. *World Wide Web*, 16, 477-496.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56-63.
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34, 185-200.
- Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78, 311-329.
- Otterbacher, J. (2009). 'Helpfulness' in online communities: a measure of message quality. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 955-964.
- Perkins, J. (2014). *Python 3 Text Processing with NLTK 3 Co*.
- Qazi, A., Syed, K. B. S., Raj, R. G., Cambria, E., Tahir, M., & Alghazzawi, D. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior*, 58, 75-81.
- Schumaker, R. P., Zhang, Y., Huang, C. N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53, 458-464.
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell system technical journal*, 30, 50-64.

- Siering, M., & Muntermann, J. (2013). What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions. In *Wirtschaftsinformatik*, 7.
- Singh, J. P., Rana, N. P., & Alkhowaiter, W. (2015). Sentiment Analysis of Products' Reviews Containing English and Hindi Texts. In *Open and Big Data Management and Innovation*, Springer International Publishing, 416-422.
- Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, 39, 1-9.
- Spool J. (2009). The Magic behind Amazon's 2.7 Billion Dollar Question. Available online at <http://www.uie.com/articles/magicbehindamazon/2009>, Accessed on 15<sup>th</sup> May 2016.
- Wan, Y. (2015). The Matthew Effect in social commerce. *Electronic Markets*, 25, 313-324.
- Wang, D., Zhu, S., & Li, T. (2013). SumView: A Web-based engine for summarizing product reviews and customer opinions. *Expert Systems with Applications*, 40, 27-33.
- Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems*, 79, 12-23.
- Wijnhoven, F., & Bloemen, O. (2014). External validity of sentiment mining reports: Can current methods identify demographic biases, event biases, and manipulation of reviews? *Decision Support Systems*, 59, 262-273.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35, 399-433.
- Wu, J., Wu, Y., Sun, J., & Yang, Z. (2013). User reviews and uncertainty assessment: A two stage model of consumers' willingness-to-pay in online markets. *Decision Support Systems*, 55, 175-185.

**Table 1**  
Summary of previous research on review helpfulness

Source	Problem statement	Solution	Data source
Ghose and Ipeirotis (2006)	Ranking product reviews	Ranks the reviews as per their expected helpfulness effect on sales.	Amazon.com
Liu, Cao, Lin, Huang, Zhou (2007)	Detecting low-quality product reviews	(1) Differentiates low-quality reviews from high-quality ones and (2) enhances the task of opinion summarization by detecting and filtering low-quality reviews.	Amazon (digital cameras)
Forman et al. (2008)	Analysis of (1) willingness to disclose identity-descriptive information about reviewers and (2) whether reviewer identity disclosure affects product sales	Identity-relevant information about reviewers shapes community members' judgment of products and reviews.	Amazon.com (books)
Danescu-Niculescu-Mizil et al., (2009)	Analyzing and modelling opinion evaluation	New correlation between proportion of helpful votes of reviews and deviation of review ratings. With average ratings, helpful votes are consistent.	Over four million Amazon.com book reviews
Mudambi and Schuff (2010)	Readers helpful vote ratio as ground truth, linear regression model based on the paradigm of search and experience goods from information economics	Review extremity, review depth, and product type affect the perceived helpfulness of the review. Product type moderates the effect of review depth on the helpfulness of the review.	Amazon.com
Ghose and Ipeirotis (2011)	Classification with random forests	Reviews that have a mixture of objective and highly subjective sentences are rated more helpful.	Amazon.com
Korfiatis et al. (2012)	Interplay between online review helpfulness, rating score, and qualitative characteristics of the review text as measured by readability tests.	Review readability has a greater effect on the helpfulness ratio of a review than its length.	37,221 book reviews collected from Amazon UK
Li et al. (2013)	Investigate product review helpfulness as well as its corresponding antecedents from the product review feature perspective (i.e., source- and content-based review features)	Customer-written reviews are more helpful than those written by experts. Customer-written product reviews with a low level of content abstractness yield the highest perceived review helpfulness.	Amazon.com
Lee and Shin (2014)	(a) How the quality of online product reviews affects the participants' acceptance of the reviews as well as their evaluations of the sources and (b) how such effects vary depending on the product type and the availability of reviewers' photos	Participants were asked to indicate how well the following adjectives described the reviews they had read: <i>very weak (1)–very strong (7)</i> , <i>not convincing (1)–very convincing (7)</i> , <i>not relevant (1)–very relevant (7)</i> . Participants' intention to purchase the product increases with positive high-quality reviews as opposed to low-quality ones.	Vitamins, DVDs, profile pictures from Facebook and Yahoo
Allahbakhsh et al. (2015)	Detecting the real rating scores of products as well as the trustworthiness of reviewers	A supporting framework for calculating a robust rating score for products, behavior analysis of reviewers, and trust computation for reviewers.	MovieLens 100k dataset
Qazi et al. (2016)	Building a conceptual model	Review types and concepts have a	1500

	for helpfulness by focusing on quantitative factors and qualitative aspects of reviewers	varying degree of impact on review helpfulness. Review length affected helpfulness negatively, that is, too-long reviews fail to attract the reader's attention and are less helpful than shorter and concise reviews.	customer reviews from TripAdvisor
--	--	--	-----------------------------------

**Table 2**

The variables used in our system

SL	Feature	Description
1	Polarity	The polarity of a given text represents whether the expressed opinion in that text is positive, negative, or neutral
2	Subjectivity	Subjectivity refers to whether the texts contains opinions and evaluations or not
3	Noun	The number of nouns in the review text
4	Adjective	The number of adjectives in the review text
5	Verb	The number of verbs in the review text
6	The Flesch Reading Ease formula	The Flesch Reading Ease Score. The following scale is helpful in assessing the ease of readability in a document: <ul style="list-style-type: none"> <li>• 90-100: Very Easy</li> <li>• 80-89: Easy</li> <li>• 70-79: Fairly Easy</li> <li>• 60-69: Standard</li> <li>• 50-59: Fairly Difficult</li> <li>• 30-49: Difficult</li> <li>• 0-29: Very Confusing</li> </ul>
7	Dale_Chall_RE	The Dale–Chall readability formula is a readability test that provides a numeric gauge of the comprehension difficulty that readers encounter when reading a text. It uses a list of 3000 words that groups of fourth-grade American students could reliably understand, considering any word not on that list to be difficult
8	Difficult_words	Difficult words do not belong to the list of 3000 familiar words
9	Length	Total words in the review
10	Set_length	Total unique words in the review
11	Stop_count	Total stop words in the review
12	Wrong_words	Words that are not found in Enchant English dictionary
13	One_letter_words	The number of one-letter words in the review
14	Two_letter_words	The number of two-letters words in the review
15	Longer_letter_words	The number of more than two-letter words in the review
16	Lex_diversity	The ratio of unique words to total words in the review
17	Entropy	The entropy indicates how much information is produced on average for each word in the text
18	Rating	The star rating of the product
19	Helpful_ratio	The ratio of total helpful to total viewed



Figure 1. A typical review on Amazon.in (Source: Amazon, 2016)

Customer Reviewer	Total Reviews	Helpful Votes	Percent Helpful
 K K R <a href="#">See all 22 reviews</a>	22	218	<a href="#">17%</a>
 * <a href="#">See all 22 reviews</a>	22	162	96%

Figure 2. Amazon reviewers ranking

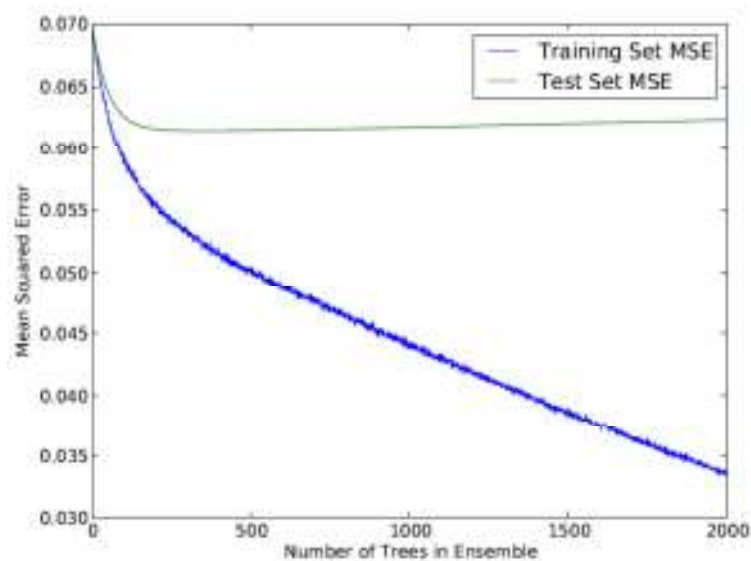


Figure 3. MSE vs. number of trees for books dataset

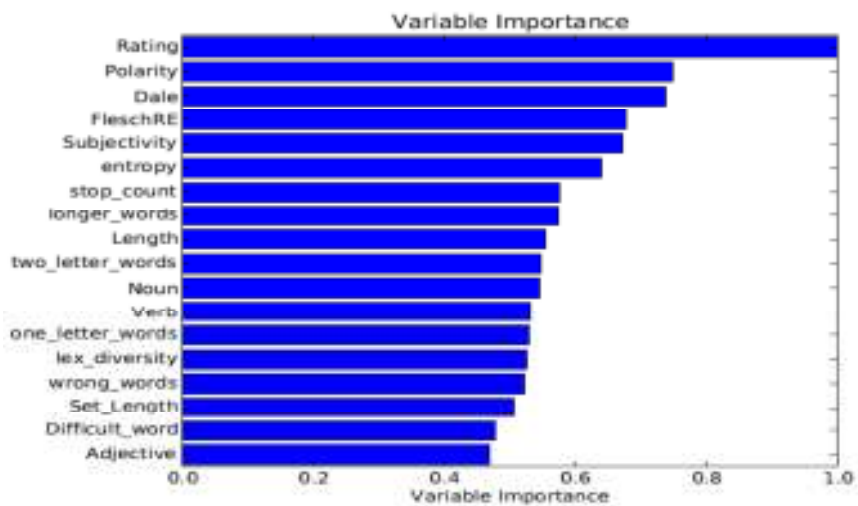


Figure 4. Variable ranking for books dataset

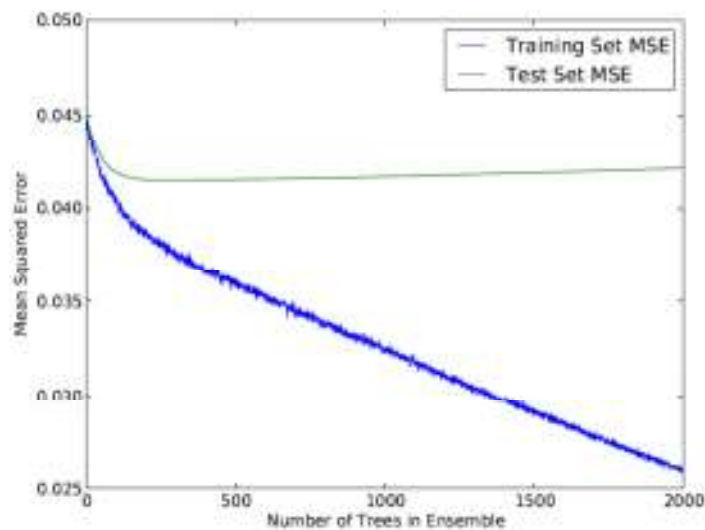


Figure 5. MSE vs. number of trees for baby products dataset

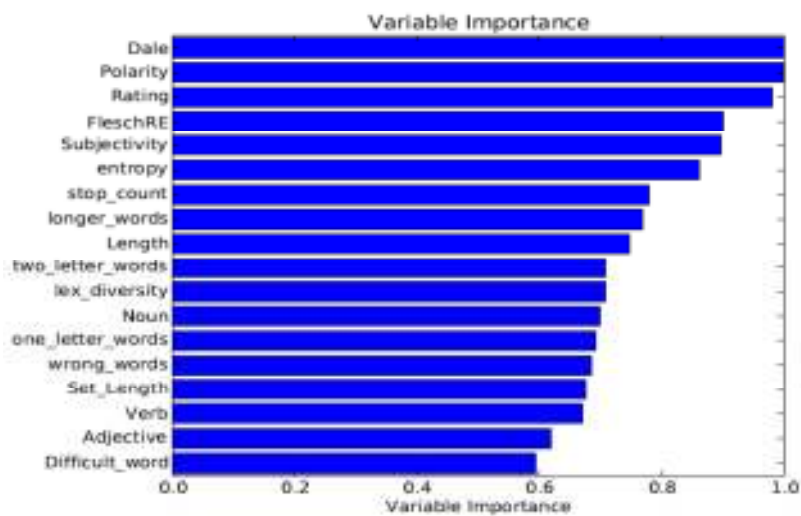


Figure 6. The variable ranking for baby products dataset

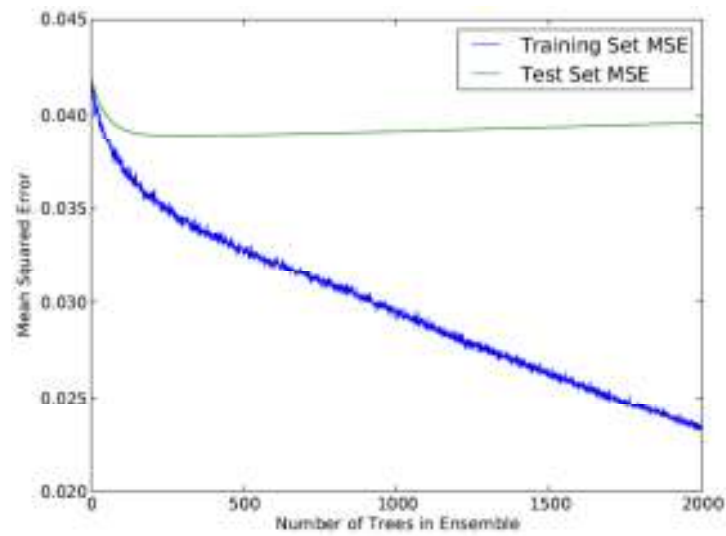


Figure 7. MSE vs. number of trees for electronic product dataset

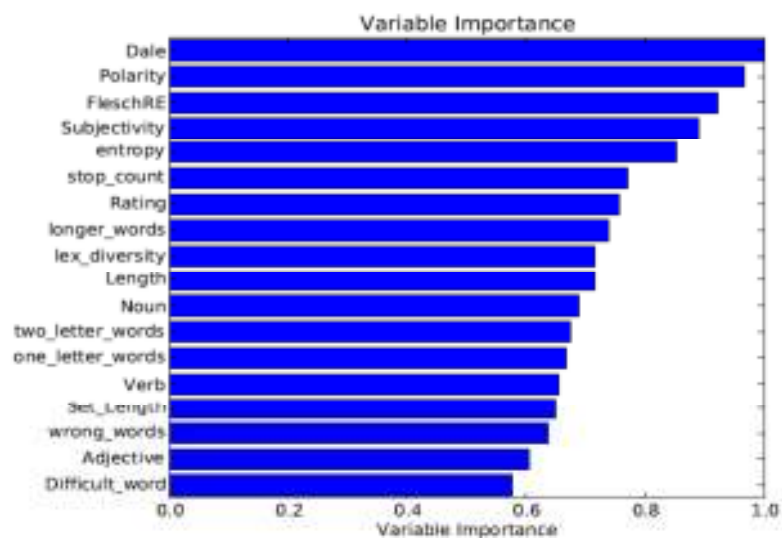


Figure 8. The variable ranking for electronic products dataset