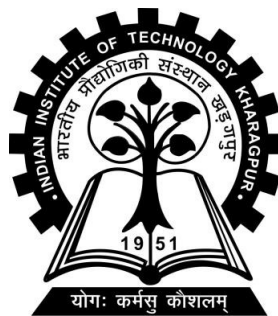


# **Review Helpfulness Prediction**

Project-I (AE47007) report submitted to  
Indian Institute of Technology, Kharagpur  
in partial fulfillment for the award of the degree of  
Bachelor of Technology  
In  
Aerospace Engineering

**By**  
**Shrey Sharma**  
**(19AE10038)**

**Under the supervision of**  
**Prof. Sujoy Bhattacharya**



**Department of Aerospace Engineering**  
**Indian Institute of Technology Kharagpur**  
**Autumn Semester, 2022-23**  
**November, 2022**

## DECLARATION

I certify that

- a. The work contained in this report was completed by me under the guidance of my supervisor.
- b. No other institute for any degree or certification has received the work.
- c. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
- d. Whenever I have used resources from other sources, I have properly acknowledged them by citing them in the thesis text and listing their information in the bibliography. Whenever appropriate, I have also secured permission from the items' rights holders.

Date: November 30, 2022

Place: Kharagpur

(Shrey Sharma)

(19AE10038)

**DEPARTMENT OF AEROSPACE ENGINEERING  
INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR  
KHARAGPUR - 721302, INDIA**



***CERTIFICATE***

This is to certify that the project report titled "**Review Helpfulness Prediction**" submitted by **Shrey Sharma (Roll No. 19AE10038)** to Indian Institute of Technology Kharagpur in partial fulfilment of the requirements for the Bachelor of Technology degree in Aerospace Engineering is a record of authentic work completed under my supervision and direction during the Autumn Semester, 2022-23.

Date: November 30, 2022  
Place: Kharagpur

Prof. Sujoy Bhattacharya  
Vinod Gupta School of Management  
Indian Institute of Technology Kharagpur

# *Abstract*

---

Name of the student: **Shrey Sharma**

Roll No. 19AE10038

Degree for which submitted: **Bachelor of Technology**

Department: **Department of Aerospace Engineering**

Thesis title: **Review Helpfulness Prediction**

Thesis supervisor: **Prof. Sujoy Bhattacharya**

Month and year of thesis submission: **November, 2022**

---

Online reviews have become the one of the most influential factor in determining the purchase behaviors and patterns of social consumers. They are a dependable type of word-of-mouth (WOM) that plays a growing role in e-commerce. However, the most difficult aspect of user-submitted product reviews is quantifying and evaluating their real usefulness. The main objective of this study is to design a technique for feature extraction that can quantify and measure each product's usefulness based on user-submitted reviews.

# *Acknowledgements*

I would like to express my sincerest gratitude to my guide, Prof. Sujoy Bhattacharya for his exceptional guidance and support, without which this project would not have been successful.

I would also like to thank the Vinod Gupta School of Management for giving me this opportunity to do this project under their esteemed institution.

Finally, I would like to thank the Department of Aerospace Engineering, IIT Kharagpur, Head of the Department Prof. Manoranjan Sinha and my Faculty Advisors Prof. Sunil Manohar Dash and Prof. Prasun Jana.

# ***1. Introduction***

Customers' ability to improve their purchasing judgments has been significantly aided by the availability of reviews posted online. Customer opinions might be swayed by reviews that were made by other customers who have previously purchased the goods. Customers have a tendency to search for the helpfulness rating on the review when they are reading the reviews. This helps them judge the reliability of the review, which in turn helps them make judgments regarding their purchases based on the review. However, when the quantity of reviews grows, it becomes practically difficult for consumers to read all of them. Additionally, customers may experience discomfort as a result of poorly written or low-quality reviews. For this reason, it is crucial for firms involved in e-commerce to locate positive reviews and present them to customers in a thoughtful manner. There is presently no automatic method for determining if a newly submitted review is useful. This is due to the fact that users are now required to vote on reviews. When one is trying to decide whether or not to make a buy, it is thus quite challenging to establish which evaluations are actually beneficial. However, this is a serious problem because the majority of customers try to create an opinion about a product before making a buy by reading various reviews of the product. People are only able to read the reviews that are directly in front of them, regardless of the total number of reviews that are available. This study attempts to establish the crucial characteristics for predicting the usefulness of a review and improve upon the work that has already been done in the field of research.

## ***2. Literature review***

Previous research has found that the utility of a review may be affected by a wide variety of elements, some of which include language qualities, the mood expressed in the review, and semantic aspects.

Reviews that are of an appropriate length, are of high readability, and include no grammatical errors have a greater chance of being seen as valuable from a linguistic point of view. Mudambi and Schuff analyse the influence that the duration of a review and the intensity of a review have on the utility of the review by analysing a large number of review datasets. According to their findings, the length of a review has a positive influence on how valuable it is, but the kind of product moderates the link between the two factors. Latent semantic analysis was utilised by Cao et al. in order to derive the meaning of reviews (LSA). They conducted research to determine which aspects of online reviews—fundamental, stylistic, and semantic—had the most impact on the utility of the reviews and discovered that the semantic aspects had the most impact. Previous research has done some investigation into the connection between the subjective nature of reviews and the value of such evaluations. In this study, we want to investigate the linguistic, semantic, and sentimental polarity features of the reviews, as well as the association between those features and the helpfulness ratings assigned to the reviews.

## ***3. Work Done: Experiments and Results***

### ***3.1 Motivation and Objective***

Online reviews have become the one of the most influential factor in determining the purchase behaviors and patterns of social consumers. They are a dependable type of word-of-mouth (WOM) that plays a growing role in e-commerce. On the other hand, the vast majority of the reviews are of a poor standard and are not worth reading. Therefore, it is extremely vital for an e-commerce firm to have the best reviews most easily available for the clients in order for them to be able to make an informed decision based on reliable reviews of the items that are being offered for sale. At the moment, the number of votes that a user who has seen these reviews casts is the sole way that can be used to classify the usefulness of reviews. However, this method is not entirely foolproof since certain reviews that include valuable information about the product are omitted from the compilation, and the consumer may not be able to access those reviews. As a result, it has turned into a need to make an automatic prediction of the usefulness of a review after it has been submitted in order to provide users with an improved experience.

The purpose of this study is to isolate the characteristics of high-quality reviews that are responsible for determining how helpful a review is. In the next section, we investigate the ways in which a review's linguistic, semantic, and sentiment polarity variables interact with one another to influence the accuracy of the helpfulness prediction.



## **3.2 Dataset**

The dataset used in this project was collected from amazon's product review data for various electronic goods. This dataset was chosen because, generally electronic goods are the most risky purchase and a consumer does proper checking of the reviews before making a purchasing decision. We collected 10000+ reviews across different products. The information of our requirement was the rating a review received, the number of helpful votes it got and the actual review body.

Numerous evaluations in our sample did not obtain a single vote of helpfulness. To counteract the skewness in the review data, we eliminated evaluations with 0% helpfulness votes and fewer than 10 words, as they were primarily consisting of filler text. Table 1 displays the relationship between the number of reviews and the number of helpful votes. To ensure sufficient observations, we divided the reviews into three categories during the analysis of the number of helpfulness votes per review: (1) reviews with zero votes, (2) reviews with between zero and five votes, and (3) reviews with five or more votes.

## ***3.3 Feature Extraction***

I utilised three distinct and combined categories of characteristics to predict helpfulness.

### ***3.3.1 Linguistic Features***

Linguistic features of a review mostly encompass the key features of reviewers' writing style that cannot be easily derived by simply browsing the review. In particular, we examine the average word length, number of stopwords (which are the commonly used words which are of no advantage such as "a", "an", "the" etc.) and number of uppercase words which a reviewer generally writes to stress on a particular point.

### ***3.3.2 Sentiment Polarity***

We measured the polarity of the review, i.e, whether the review is a positive review or a negative review to determine whether that has an effect on the helpfulness rating of a review. This was achieved using VADER (Valence Aware Dictionary and sEntiment Reasoner) module of python. When the data being analysed has not been labelled in any way, VADER is able to determine the polarity of a sentiment inside a particular body of text by determining how positive or negative it is. In order to assess the overarching sentiment of a given body of text, it consults a lexicon of words associated with particular emotions. An example of how it works is shown below:

Word	Sentiment rating
Tragedy	-3.4
Rejoiced	2
Insane	-1.7
Disaster	-3.1
Great	3.1

*Fig. Each word has a Valence Rating*

It not only provides information on the positivity and negativity score, but it also provides information regarding the degree to which a sentiment is favourable or negative. It provides a sentiment score that ranges from -1 to 1, with -1 being the least positive and 1 being the most favourable. The calculation for the sentiment score involves adding together the scores assigned to each word in the VADER list that is included in the phrase. It is important to keep in mind that the sentiment score of individual words ranges from -4 to 4, but the sentiment score returned for a whole sentence ranges from -1 to 1. The sentiment score of a sentence is calculated by adding the sentiment scores of all the words in the sentence that carry emotional weight; this combined score is then normalised so that it falls somewhere between -1 and 1. The normalization technique used is:

$$\frac{s}{\sqrt{s + \alpha}}$$

where, 's' is the sum of the sentiment scores of the words and alpha is the normalization parameter which is set to 15.

VADER, works best on short documents, like tweets and sentences, not on large documents.

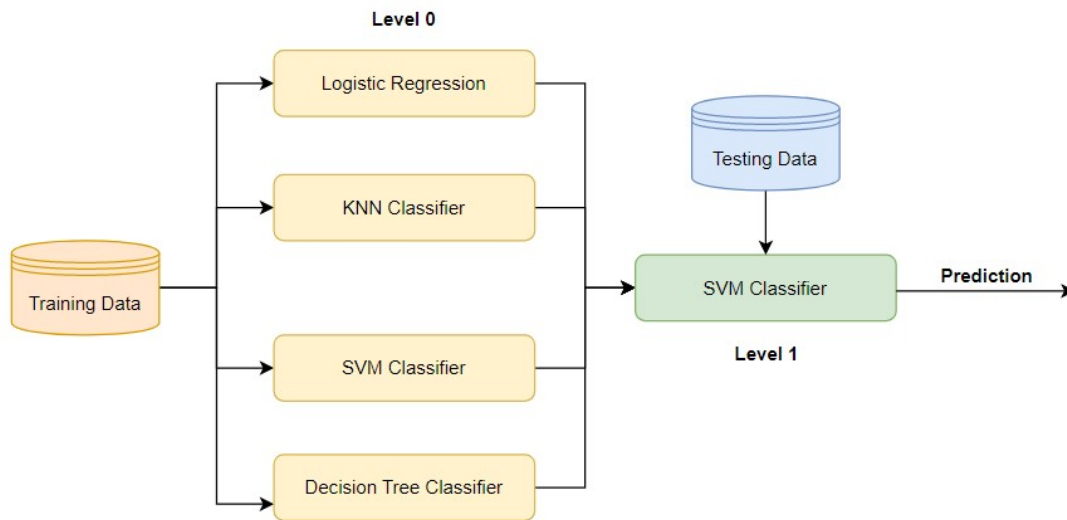
### **3.3.3 Semantic Features**

While analysing the semantic features of a review, we attempt to extract insightful information, such as context, emotions, and sentiment, to emphasise their significance in a review's usefulness. We attempt to parse the meaning of reviews with the help of Word2Vec embeddings. Word2Vec is a group of related models that are used to produce word embedding. These are shallow, two-layer neural networks that are trained to reconstruct linguistic contexts of words. It can be obtained using two methods: Skip Gram and Common Bag of Words (CBOW). In this research, we used the skip-gram model of word2vec to extract each word as a vector and then used techniques such as cosine similarity to determine the similarity between distinct words.

## **3.4 Model description**

For training the model, we selected conventional classification models including Logistic Regression, KNN Classifier, SVM Classifier, and Decision Tree Classifier. Finally, we used a stacking ensemble classifier where all the above models were used as the Base-models and SVM Classifier as the Meta-model.

Here, the meta-model is trained using the base models' predictions. In other words, data that was not used to train the base models is fed to the base models, predictions are formed, and these predictions, along with the predicted outputs, provide the input and output pairs of the training dataset needed to fit the meta-model. The sketch of the stacking ensemble model used is given below:



**Fig. Stacking Ensemble Model**

Now, in order to examine how different characteristics of reviews influence the number of votes they receive, we sampled the different features to generate training datasets for the model:

- Linguistic + Sentiment + Semantic features
- Semantic + Sentiment features
- Linguistic + Sentiment features

## 3.5 Results

The table below provides a summary of the empirical results of comparing the various models on the various sampled datasets. We discovered that the stacked classifier model, which combines all linguistic, sentiment, and semantic features, has the highest accuracy and is therefore the most effective model. These results demonstrate that combining semantic characteristics with linguistic and sentiment polarity significantly improves the model's performance. These results indicate that semantic characteristics play

a significant role in determining the number of helpfulness votes a review receives. The model that incorporated linguistic and sentiment feature characteristics performed second best overall. The third model, which incorporates semantic and sentimental characteristics. These results indicate that the combination of all feature classes is essential for accurately classifying reviews as helpful.

	Models				
Datasets	Logistic Regression	KNN Classifier	Decision Tree Classifier	SVM Classifier	Stacking Classifier
Linguistic + Semantic + Sentiment	0.6244	0.6475	0.5821	0.6588	<b>0.7012</b>
Semantic + Sentiment	0.5788	0.5745	0.5266	0.5895	0.6711
Linguistic + Sentiment	0.6477	0.6022	0.6150	0.6277	0.6286

**Table:** Accuracy of different models on the different feature datasets

## ***4. Discussion and Future Work***

In this project, we sought to identify the crucial characteristics that determine the usefulness of a user review. We discovered that linguistic characteristics such as the number of stopwords and the average number of words per review play a significant role. If the number of stopwords in a review is high, it is logically classified as unhelpful, as these reviews have little useful information to contribute. We also considered the sentiment polarity of each review, categorising them as positive or negative based on their VADER score. We discovered that the extremeness of the review, whether positive or negative, was a significant factor in classifying reviews in a useful manner. Using the word2vec embedding vectors, we also investigated the semantic characteristics of a review. The results demonstrated that semantic characteristics play an important role in the useful classification of a review. There are also several limitations to these results. Although we discovered that semantic characteristics play a significant role in encouraging helpfulness votes, we did not investigate them in depth. The sentiment polarity analysis is also a limitation of our study. We did not investigate precisely how much the positivity or negativity of a review affects its usefulness. These limitations necessitate further study. It would be intriguing to delve deeper into the semantic characteristics of the reviews and determine which characteristics garner more votes of helpfulness. We may also observe that the extent of a review's positive or negative sentiment influences the helpful votes. We can also attempt to enhance our model using neural network techniques such as BERT.

## 5. 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., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511-521.
4. Haque, M. E., Tozal, M. E., & Islam, A. (2018, August). Helpfulness prediction of online product reviews. In *Proceedings of the ACM Symposium on Document Engineering 2018* (pp. 1-4).
5. <https://towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53>
6. <https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/>



# 19AE10038 - BTP REPORT

## ORIGINALITY REPORT

13%

SIMILARITY INDEX

11%

INTERNET SOURCES

10%

PUBLICATIONS

10%

STUDENT PAPERS

## PRIMARY SOURCES

1	home.gwu.edu Internet Source	5%
2	machinelearningmastery.com Internet Source	2%
3	piocalderon.github.io Internet Source	2%
4	oa.upm.es Internet Source	1%
5	"Youtube Video Ranking: A NLP based System", International Journal of Recent Technology and Engineering, 2019 Publication	1%
6	Cao, Q.. "Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach", Decision Support Systems, 201101 Publication	1%
7	Submitted to University of Sydney Student Paper	1%

8	Submitted to University of Computer Studies	1 %
Student Paper		

9	Jiwan Seo, Karam Yoo, Seungjin Choi, Yura Alex Kim, Sangyong Han. "The latent learning model to derive semantic relations of words from unstructured text data in social media", Multimedia Tools and Applications, 2018	<1 %
Publication		

10	journal.growthguide.co.in	<1 %
Internet Source		

11	aisel.aisnet.org	<1 %
Internet Source		

Exclude quotes    On

Exclude matches    Off

Exclude bibliography    On