Prediction of Customer Review's Helpfulness Based on Feature Engineering Driven Deep Learning Model

Surya Prakash Sharma, Department of Computer Science and Engineering, Dr. APJ Abdul Kalam Technical University, Lucknow, India

Laxman Singh, Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida, India*

Rajdev Tiwari, Edunix India Pvt. Ltd., Noida, India

ABSTRACT

Online consumer reviews play a pivotal role in boosting online shopping. After Covid-19, the e-commerce industry has been grown exponentially. The e-commerce industry is greatly impacted by the online customer reviews, and a lot of work has been done in this regard to identify the usefulness of reviews for purchasing online products. In this proposed work, predicting helpfulness is taken as binary classification problem to identify the helpfulness of a review in context to structural, sentimental, and voting feature sets. In this study, the authors implemented various leading ML algorithms such as KNN, LR, GNB, LDA and CNN. In comparison to these algorithms and other existing state of art methods, CNN yielded better classification results, achieving highest accuracy of 95.27%. Besides, the performance of these models was also assessed in terms of precision, recall, F1 score, etc. The results shown in this paper demonstrate that proposed model will help the producers or service providers to improve and grow their business.

KEYWORDS

Binary Classification, CNN, Machine Learning, Online Reviews, Reviews Feature Set, Reviews Helpfulness

1. INTRODUCTION

Review helpfulness is the part of business intelligence (BI) and plays a pivotal role for the e-commerce business to populate their sites with number of genuine reviews to assist customers by their products and services. Google provides cumulative rating of a product based on the reviews and rating that it receives from various sources like pbtech, eBay, and Samsung. Online customers mainly like to read the online reviews related to that particular product, which they want to buy. Nowadays, plenty of reviews are available to help the online customers in deciding them about the right product. These days' e-commerce websites try to discern the usefulness of reviews by conducting an online survey or through telephonically. Online review sites such as Yelp, Amazon, etc., offers millions of customer reviews, which might have greater impact on the market trends as well as on buying decisions of

DOI: 10.4018/IJSI.315734 *Corresponding Author

many potential customers (Guo et al., 2020). Aa per Murphy et al., 2020, about 86% of customers used to read online reviews and shows a deeper trust in them. By year 2020, more than 200 million reviews were published on Yelp. The review helpfulness provides the insight about the subjectivity and quality of the product (Li et al., 2019; Filieri et al., 2018). In general, helpful votes are treated as a true indicator of review reliability (Huang et al., 2015). The prime challenges of customers and businesses firms are to get the benefit from such reviews which are available in bulk quantity and are of inconsistent nature. Out of the several aspects of the reviews, helpfulness is taken as key feature and researched widely. Review helpfulness is computed by calculating the number of helpful votes divided by the total number of votes (Bilal et al., 2021). Figure 1 shows how Amazon.com gathers helpful votes of the reviews from their readers (Park et al., 2018). A lot of work in the direction of predicting helpfulness has already been carried out as of now, details of which are given in the subsequent section.

2. LITERATURE REVIEW

Online product reviews are suggested as a valuable tool for promoting products, as well as for collecting consumer feedback and boosting sales (Chua et al., 2014; Forman et al., 2008; Hu et al., 2008). In literature (Gang et al., 2008), there have been shown the direct relationship between product ratings and sales. For example, online movie reviews and ratings have significant impact on box office revenues (Lee et al., 2018). Similarly, online book reviews have positive impact on book sales (Chevalier et al., 2006). Authors in (Zhang et al., 2006), put forward a regression model to predict the utility of product reviews. In these studies, the authors utilized lexical similarity, and syntactic terms based on Part-of-Speech (POS) as features.

Review helpfulness is considered subjective assessment by numerous authors with respect to review quality that indicate review true diagnostics (Huang et al., 2013). Helpfulness is observed as the perceived cumulative value of the information encased in a review (Cao et al., 2011). Over the past two decades, review helpfulness are evaluated using star rating, reviewer credibility, and product price and its types (Otterbacher et al., 2009; Mudambi et al., 2010). Product, review metadata, and review characteristics are some common features authors (Lee et al., 2014; Wang et al., 2018), to design a multilayer perceptron model for helpfulness prediction. In their work, authors mainly focused on the improvement of helpfulness prediction by using neural network model over the linear regression models (Krishnamoorthy, 2015).

With huge availability of online reviews, it may be quite cumbersome for customers to differentiate between helpful & non-helpful reviews. Helpful votes were initially introduced by Amazon (Wan, 2015).

Ref. (Liu et al., 2015), presents a helpfulness prediction model for travel goods websites. As features for the prediction problem, they have employed the reviewer's experience, the reviews' writing style, and the reviews' timeliness. Researchers have been frequently utilizing regression models to

Figure 1. A typical review on Amazon (Source: Amazon, 2022)

evaluate various types of textual and non-textual reviews to get the details about the factors that affect the review helpfulness (Krishnamoorthy, 2015). In (Mudhambi et al., 2010), authors suggested three hypotheses to discern the review characteristics based on the data collected from Amazon.com. These hypotheses have been proven beneficial for potential customers. After testing the theory, impact of review extremity was found to be different based on the different product types. In contrast to a conventional review helpfulness, here authors determined helpfulness using the confidence interval and the helpfulness distribution data. In their study, for the purpose of experimentally proving the method's usefulness, the authors employed both artificial and actual datasets (Krishnamoorthy, 2015). Table 1 list out the summery of the key findings of different state of art methods with regards to reviews helpfulness.

In this study, other important qualities, such as subjectivity, readability, and meta-data aspects (S.-M. et al., 2006; Cao et al., 2007; Ghose et al., 2011), which were empirically demonstrated to be superior to other predictors before, were excluded. Recently, Quaschning et al., 2015, investigated the connection between valence consistency and review helpfulness to study the effect of adjacent. In one of the study, Chen et al., 2011, divided product reviews into low, medium, high, and spam that were then analyzed for review helpfulness. Subsequently, a model was created by integrating readability, subjectivity, and information related factors to predict low- or high-quality reviews. In ref. (Zhang et al., 2014), authors conducted numerous experiments on synthetic and real datasets to analyze the connection between helpfulness ratio, helpfulness distribution and confidence interval. The proposal to extract fresh linguistic features from the text reviews was utilized in (Krishnamoorthy, 2015), that were used to get better review helpfulness prediction result in term of accuracy and hybrid set of features. In ref. (Huang et al., 2015), authors inspect the importance of quantitative as well as qualitative features such as reviewer's impact, experience and cumulative helpfulness of reviews and reviewers, respectively. The authors showed the impact of threshold value on word count and demonstrated the varying effect of reviewer experience on helpfulness prediction. Ref. (Malik et al., 2017), used regression model to explore the importance of textual and non-textual features in helpfulness prediction. In their study, authors classified the products into two categories; first was experience based product and other was search based products. Besides prediction of review helpfulness, machine learning and deep learning techniques have also been widely used for the different types of applications such as cancer detection (Dafni et al., 2022; Srinivas et al., 2022; Ramanan et al., 2022), plant disease detection (Deepkiran et al., 2022), unstructured road detection

Table 1. Summery of the key findings on the helpfulness of reviews in current literature

References	Data source	Problem statement	Contribution
(Mudambi et al., 2010)	Amazon.com	Used linear regression model for helpfulness prediction of experienced goods based on the paradigm of information economics.	Review depth, review extremity, and product type tends to affect the perceived review helpfulness. Review depth falls positive impact on the review helpfulness.
(Li et al., 2013)	Amazon.com	Examined product review helpfulness as well as its corresponding source- and content-based review features.	Customer-written product reviews were found more helpful in comparison to those done by experts.
(Cao et al., 2011)	CNET	This work mainly focuses on increasing the number of votes for helpfulness.	The semantic features are more influential than other features in affecting how many helpfulness votes reviews receive.
(Krishnamoorthy, 2015)	Amazon.com	Built a predictive model for examination of the factors influencing the helpfulness of online reviews.	For experienced and search goods, linguistic category features (Verb, adjective etc.) are considered more impactful.

and classification (Alam et al., 2021), fault diagnosis in insulators (Singh et al., 2021) as well as for solving the robotic path planning problems (Kumar et al., 2021).

A literature summary on reviews helpfulness prediction based on classification problems is given in Table 2.

Krishnamoorthy et al., 2015 built a predictive model to examine the factors that have a potentially higher impact on the helpfulness of online reviews. A linguistic feature (LF) value was created using the specified model to extract linguistic aspects including adjectives, state verbs, and action verb features. Other review metadata like 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, etc.) were also included in their model to predict helpfulness.

Based on the features of customer evaluations, a deep neural network was proposed in this article. The following are the key contributions:

1. In this study, unique features related to product reviews are taken from Amazon datasets and separated into three categories, including voting, structural and sentimental aspects. The impact of feature sets is also investigated for review usefulness.
2. Furthermore, the CNN model is customized and trained using manually created review features to create an efficient prediction model for predicting the review usefulness of a particular product.
3. The result of the proposed model is compared with the other existing state of art methods using the same prediction parameters.

The remaining portions of the article are arranged as follows: The Methodology Section presents the existing work related to helpfulness prediction on online customer reviews; Feature Engineering manually extracts the review features and divides them into subsets; Experiments Section illustrates and compares the experimental results with existing and past best models; and Conclusion Section contains the concluding remark.

Table 2. A summary of current research on the classification issue of reviews' helpfulness

Reference	Data Sources	Helpfulness	Feature			Prob. Type		ML Algorithms	Predicted Result
			R_c	P	R	C	R_{eg}		
(Singh et al., 2017)	Amazon.com	ratio	✓	×	×	×	✓	GBT	MSE
(Krishnamoorthy, 2015)	Amazon.com		×	×		✓	×	NB, SVM, RandF	Accuracy, F1 score
(Ghose et al., 2011)	Amazon.com	ratio	✓	✓	✓	✓	×	RandF, SVM	Accuracy, AUC
(Lee et al., 2018)	TripAdvisor	ratio	✓	×	✓	✓	✓	LGR, DT, RandF, SVM	Accuracy, F1 score, AUC
(Malik et al., 2017)	Amazon.com	ratio	✓	×	×	✓	×	NB, RDA SVM, RandF, Pruned C4.5, DNN	Accuracy, F1 score
(Wang et al., 2018)	Amazon.com	ratio	✓	✓	✓	×	✓	LNR, MLP	MSE

Note: R_c (Review Content), P (Product), R (Reviewer), ML (Machine Learning),C (Classification), R_{eg} (Regression), (Gradient Boosted Tree), LNR (Linear Regression), RandF (Random Forest), GBT (Gradient-Boosted Trees), DNN (Deep Neural Network), *LGR* (Graphical model for Local Gaussian Regression).

3. RESEARCH METHODOLOGY

In this study, we formulated the classification model for user-generated online reviews' helpfulness prediction. Three feature sets with different numbers of features have been used as an input to build the review helpfulness prediction model. In this experiment, the most recent baseline attributes (Krishnamoorthy, 2015; Malik et al., 2017; Chau et al., 2018; Gang et al., 2018: Park et al., 2018; Li et al., 2019; Bilal et al., 2021), have been considered for building the CNN based prediction model.

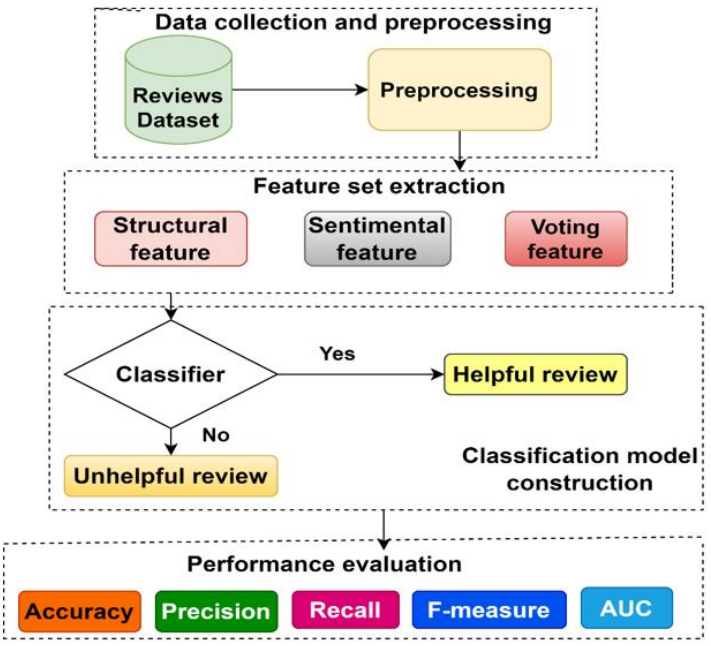
As indicated in Figure 2, the framework used in this study include five steps; (i) data gathering, (ii) preprocessing, (iii) issue characterization, (iv) feature set extraction, and (v) model creation. The review information used in this study have been obtained from the data source available online freely on Amazon.com. The available dataset has three parts: a corpus of reviews, social data related to those reviews, and reviewer data.

4. DATA COLLECTION AND PREPROCESSING

The data set used to conduct the experiment was initially taken from Amazon.com between May 1996 and July 2014 and was accessible at http://jmcauley.ucsd.edu/data/amazon/. There are over 83 million unique reviews on Amazon, which cover 24 major product categories (Alsmadi et al., 2020).

Review information was collected for cellphones and accessories, with raw data sizes of 138 MB. Products like phones and accessories comes under the category of search goods, making it very simple to learn about a product's quality before interacting with it (Park et al., 2018). The nine attributes reviewerID, asin, reviewerName, helpful, reviewText, overall, summary, unixReviewTime, and reviewTime are all present in each entry in the dataset. The details of each characteristic are shown below:

Figure 2. Framework of research methodology



1. **reviewerID:** The alphanumeric code that Amazon gives to its reviewers.
2. **asin (amazon standard identification Number):** The alphanumeric product ID given to a goods by Amazon.
3. **reviewerName:** The reviewer's Amazon.com name is listed in the review.
4. **helpful:** The proportion of users who considered the review useful, for example [10/15].
5. **reviewText:** The reviewer's submitted review content.
6. **overall:** The product's star rating value.
7. **summary:** The review's summary.
8. **unixReviewTime:** The time of the review was published (as Unix time).
9. **reviewTime:** The review's creation date and time.

4.1 Preprocessing

We utilized real-world review dataset that to run several experiments to measure the effectiveness of different applied ML models under study for categorization of helpfulness review problems. On the dataset, a data cleaning method is used to eliminate redundant reviews and enhance the effectiveness of our research (Malik et al., 2017).

The steps in data cleaning process are:

1. The first step deals with the identification and removal of duplicate reviews from datasets.
2. Second step involves the removal of blank text in reviews from the dataset.
3. Third step involves the filtration of reviews with a high percentage of votes for getting better classification results.

As a result, only reviews with at least 10 total votes are considered. After the preprocessing procedure, dataset contains 8775 customer reviews.

4.2 Feature Set Extraction

In this section, the initially, the classification problem is followed by the characteristic set to be employed in our model. The characteristics set that was employed in our prediction model is then described.

Suppose $R=\{ \alpha^n, \beta^n\}$ where $\alpha^n =\{ \alpha_{1,} \alpha_{2,} \ldots\ldots \alpha_n \}$ represents the features of n numbers of reviews and $\beta^n \in \{0,1\}$ represent the levels of helpfulness. '1' indicate review is helpful, and '0' indicate review is unhelpful. Features have a significant role in how well a categorization model performs. In our problem statement, each x_i represents a feature vector has k number of features where $k \in \mathbf{+I}$ (universal set of positive integer number), is composed of three disjoint feature set $\mathbf{[S_t, S_e, V]}$ where $\mathbf{S_t}$, $\mathbf{S_e}$, $\mathbf{V}$ describes the structural, Sentimental and voting feature set reviews respectively. The combination of features set $\mathbf{[S_t, S_e, V]}$ form the overall feature matrix, α^n.

This study uses 17 reviews feature based on the literature (Krishnamoorthy, 2015; Malik et al., 2017; Chau et al., 2018), that grouped into three disjoint feature sets. While the voting features set is derived from the helpful and unhelpful reviews used in reviews dataset. Following are more detailed descriptions of each of the features used in our study.

4.2.1 Structural Feature Set (S_t)

Features related to the organization of text reviews are included in this feature set. We employed 10 structural characteristics in this investigation. Most structural aspects of a review make use of various content frequency measurements. Utilizing structural elements serves the purpose of capturing the significance of specific words, phrases, and passages of the text. The following list of structural traits was used in our experiments:

Structural feature set (S_t) = {Char_Len, nWord, nSent, WPS, DFW, UW, Rating, Avg_W_Len, 1Word, 2Word, Long_Word} (1)

- **Char_Len:** Character count for the entire review content.
- **nWord:** Overall number of words in a single review test.
- **nSen:** Overall number of sentences in a review.
- **WPS:** Average number of words in a sentence.
- **DFW:** Difficult words used in review that is beyond the list of 3000 familiar words.
- **UW:** Words that are distinctive in a review.
- **Avg_W_Len:** Average word length in a review.
- **1Word:** Total number of one length word in a review.
- **2Word:** Total number of two length word in a review.
- **Long_Word:** Total number of more than two length word in a review.

4.1.2 Sentimental Feature Set (S_e)

In this experiment, four review features are used in sentimental feature set:

(S_e) = {Neg, Pos, Neu, Compound} (2)

- **Neg:** Negative sentiment score is measured in the scale from - 4 to 0.
- **Pos:** Positive sentiment score is measured in the scale from 0 to + 4.
- **Neu:** Natural sentiment represent by mid-point 0.
- **Compound:** It is normalized score between – 1 to + 1 of the sums of Neg, Pos and Neu.

Using the VADER tool for vocabulary- and rule-based sentiment analysis, numerical representations of the polarity ratings from reviews were generated (Valence Aware Dictionary and sEntiment Reasoner) (Hutto et al., 2014), using python library. These sentiment scores were obtained by VADER by converting lexical features into emotion scores and adding the four characteristics of neutrality score, positivity score, negativity score, and compound score to our collection of sentimental features. The total score is summarized by the compound score.

4.1.3 Voting Feature Set (V)

Voting feature set (V) includes three features, which are given as follows:

- **Total:** Overall number of votes received.
- **Vote:** Overall number of helpful votes received.
- **Helpfulness:** Ratio of total helpful vote divided by total number of votes in the range of [0, 1].

In addition to literature (Krishnamoorthy, 2015; Malik et al., 2017; Bilal et al., 2021), researcher labeled the review helpfulness in to two classes ‘1’ and ‘0’, by use of some threshold value (Φ).

Suppose $\mathbf{H_i}$ represents the helpfulness of a review i, then for classification task:

$$\mathbf{H_i} = \begin{cases} 1(helpful), & if\ helpful\ ratio \geq \Phi \\ 0(unhelpful), & otherwise \end{cases} \quad (3)$$

where the important parameter in our classification task is the helpfulness threshold value Φ . For our experiment default value of Φ is 0.60 is based on past research in literature (Ghose et al., 2011; Hong et al., 2012; Malik et al., 2017).

5. BUILDING THE CLASSIFICATION MODEL

We used the feature matrix of size N x F in the feature engineering process to build the classification model, where n is the total number of reviews and F is the number of features. Here, value of F is taken as 17. In this work, we used well-known machine learning algorithms to categories a dataset of product reviews as either helpful or unhelpful based on cutting-edge feature sets. Machine learning methods were used to create the usefulness predictive model as 1. K-nearest neighbor (KNN) 2. Linear regression (LR) 3. Gaussian Naive Bays (GNB) 4. Convolution Neural Network (CNN). Review dataset was divided into two parts. First part utilized as training set and second part as testing set. Datasets split into the ration of 80:20. Training set kept 80% of the dataset, and this dataset was used for trained the model using popular machine learning algorithms. The remaining 20 percent was used to assess the system's performance. We utilize the fundamental programming language Python 3.6 to create and test each model.

Five assessment measures are used in addition to this study to assess the effectiveness of classification model performance. Precision, Recall, Accuracy, F1 Score, and AUC are some of the assessment measures. These assessment metrics are represented mathematically as:

$$\text{Precision} = \frac{Tp}{Tp + Fp} \tag{4}$$

$$\text{Recall} = \frac{Tp}{Tp + Fn} \tag{5}$$

$$\text{Accuracy} = \frac{Tp + Fp}{Tp + Tn + Fp + Fn} \tag{6}$$

$$\text{F1} = 2* \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

$$\text{Area under curve (AUC)} = \int_{b}^{a} f\left(x\right) dx \tag{8}$$

5.1 Convolutional Neural Network Architecture

In the current study, we implemented CNN based deep neural network model for the binary classification problem of review helpfulness data. The undertaken model was implemented in python using Keras library in addition, Tensorflow and Theano were used as a backend. CNN has

been identified as the best solution for binary classification problems due to its self-feature learning capabilities and outstanding predicted results.

The proposed model used in our study (as shown in Figure 3) consist of two 1-dimensional convolution layer (Conv1D). Each convolution layer (Conv1D) has a pooling layer (Pool), batch normalization, and a rectified linear unit (ReLU) activation function. It also has drop out, sigmoid, fully connected (fc), and classification output layers. Besides "Binary cross entropy" was chosen as the loss function for the output layer. In this model, we sequentially passed the feature information of 17 neurons from the feature extraction stage to the convolution input layer. As discussed above, N * 17 is the size of the feature matrix that comprises of three feature sets. There are a total of 12864 neurons applied on the second convolutional layer (Conv1D 2) and 64 neurons were applied on the first convolutional layer (Conv1D 1). Subsequently, output layer had only single neuron to display the output. The result of output layer is bounded by (0, 1). The result of output layer is divided into two classes, helpful and unhelpful. Helpful is indicated by the value 1 and unhelpful is indicated by the value 0 for the online product review dataset.

The other tuning parameters of CNN model are described in Table 3.

6. EXPERIMENTAL RESULTS

In this work, we conducted a various experiments utilizing different machine learning models on suggested feature sets and compared how well these models performed on user-generated reviews using various evaluation matrices.

Figure 3. Structure of CNN model

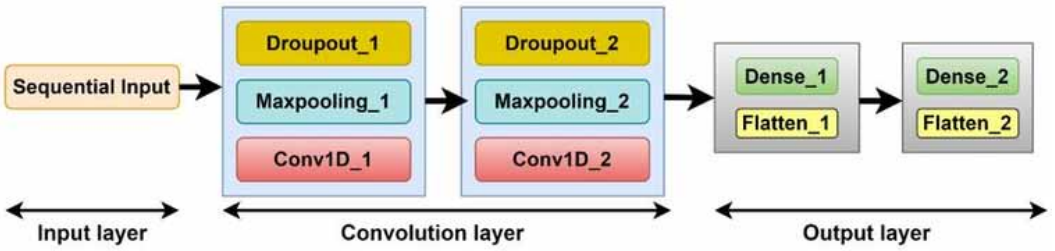


Table 3. Tuning parameters of proposed model

S.No.	Parameters	Value
01	Optimizer	Adam
02	Learning rate	0.0005
03	Beta_1	0.9
04	Beta_2	0.999
05	Epsilon	0.00000001
06	Epochs	24
07	Loss function	Binary Cross entropy
08	Verbose	1
09	Batch Size	20

6.1 Performance Analysis of ML Models

The machine learning methods employed in this study have shown to be highly effective in predicting how helpful reviews would be. Four prediction models for the usefulness of user-generated reviews are created using a 10-fold cross-validation method on the specified feature set. The four machine learning (ML) methods viz., Logistic Regression (LR), K-Nearest Neighbors (KNN), Gaussian Nave Bayes (GNB), and Convolutional Neural Network (CNN) are used to create predictive models. The experiment predicts the utility of proposed feature sets using product datasets. Precision, Recall, Accuracy, F1 score, and AUC are some of the assessment measures used to assess the usefulness of predictive models and compare their performance. Table 4 presents comparison findings for different prediction models and highlights the predictive model with the best outcomes.

As can be seen in Table 4, CNN predictive model achieved good results in terms of F1 and accuracy as compared to other competitive algorithms. The model obtained F1 of 96.0% and accuracy of 95.27%. Based on the results shown in Table 4, CNN model was found to be most efficient model that yielded comparatively promising and competitive results than the other state of art algorithms.

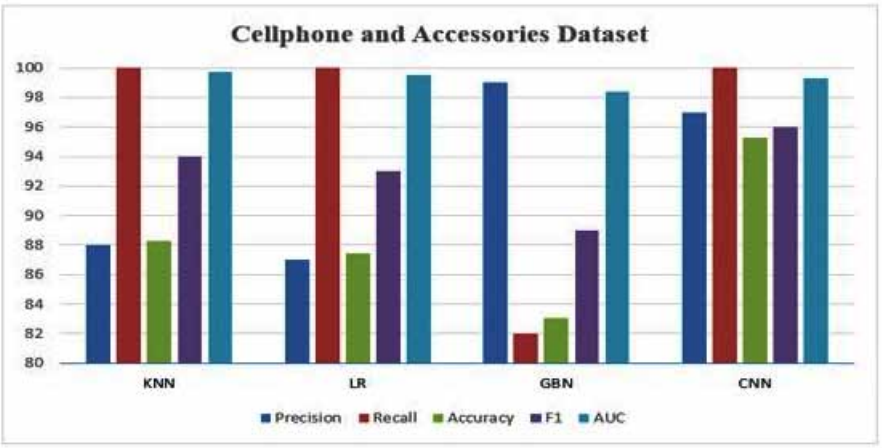
In comparison to other models, the Gaussian naive bays (GNB) classification model had the lowest performance for all the assessment parameters except precision. For review dataset, the minimal accuracy of GNB model was 83.08 percent. Figure 4 presents predictive performance in terms of visualization results, for all the four models.

Based on the above discussion, we can conclude that CNN model generated promising output for review and usefulness prediction. The outcomes demonstrate the proposed model for review classification problem.

Table 4. Predictive model performance using 10-fold cross-validation

Dataset	Model	Precision (%)	Recall (%)	Accuracy (%)	F1 (%)	AUC
Cellphones and accessories	KNN	88	100	88.31	94	99.7
	LR	87	100	87.4	93	99.5
	GNB	99	82	83.08	89	98.4
	CNN	97	100	95.27	96	99.3

Figure 4. Predictive model performance in percentage (%) of ML algorithms for dataset



6.2 Comparison of CNN Model Based on Different Evaluation Matrices

In Table 4, results of all the models under study are presented. The results shows that the CNN model shows the best classification performance for user-generated review helpfulness dataset. Table 3 list out the tuning parameters used in our model. The model completes its evaluation in 119 seconds 24 epochs. Figure 5 shows the behavior of our proposed model's training and validation after 24 epochs on the review's dataset. The model obtained the accuracy of 93.09% and 95.27% for training and validation phases, respectively, while obtained the result of 16.87 and 17.96 percent in terms of loss for training and validation phases. Besides, these hybrid feature sets demonstrate the precision, recall, accuracy, F1, and AUC with score of 97.90%, 100%, 95.27%, 96.00%, and 99.30, respectively.

Figure 6 presents the precision recall curve for reviews the dataset to comprehend the impact of these hybrid feature sets on classification results. In the result, our suggested model predicts the AUC value to be 99.30 percentage which is comparable to the others ML models.

The outcomes further demonstrate the predictive ability of ML algorithms and found CNN to be an effecive predictive model for user-generated review helpfulness prediction. Here also, the poorest performance was observed in case of GNB model. The experimental results reveals that CNN provides better prediction performance across all assessment parameters compared to KNN, LR, and GNB.

Figure 5. Accuracy and loss of CNN predictive model for Cellphone and Accessories Dataset

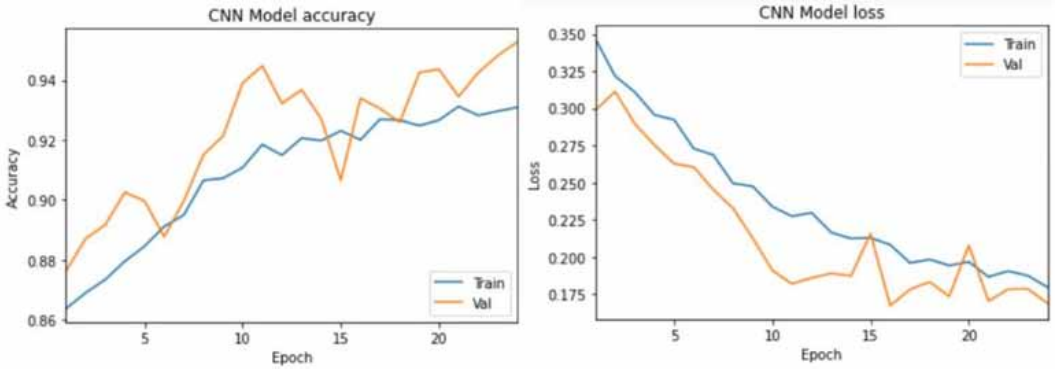
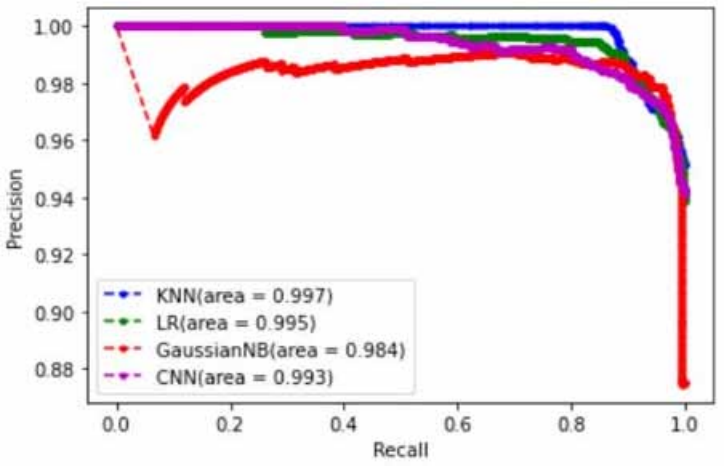


Figure 6. Precision recall curve for AUC for Cellphone and Accessories dataset



6.3 Comparison With Best Previous Models

In Table 5, the performance of our suggested model is compared with the other recent state of art models reported in the literature. The results shown in Table 5 reveals that our model outperforms over the other recently developed state of art models. As per the results shown in Table 5, CNN yielded the improved results in terms of precision by 27.88% over the (Bilal et al., 2021) methods. This illustrate that the proposed approach has good prediction capability for evaluation of both helpful and unhelpful reviews. It was observed that our method scored F1 score of 96.00, which outperformed contemporary state of art models (Krishnamoorthy, 2015; Malik et al. 2017; Bilal et al. 2021).

Prior models have not considered the precision recall curve for study, which is close to 1 in our experiment and provides a strong indicator of our suggested model's classification ability. (Bilal et al. 2021), used AUC (area under the curve) to evaluate the model performance and obtained the value of 0.773, while our model obtained the value of 0.993 which is far better than the performance of the existing model. In terms of accuracy too, our model produced better results than those given in the literature.

7. CONCLUSION

In this study, various machine learning models viz., CNN, K-nearest neighbor (KNN), Linear regression (LR), Gaussian Naive Bays (GNB) are implemented for classification of helpful and unhelpful reviews on Amazon datasets. To train the models, we used combination of three set of features namely voting, structural and sentimental characteristics, which are derived from previous studies. Out of all the considered models, CNN scored highest with the accuracy and precision of 95.27% and 97%. The results of CNN model were compared against various other competitive machine learning algorithms such as KNN, Linear regression LR, Gaussian Naive Bays GNB. The numerical and graphical results of the aforementioned models demonstrate that CNN model has the potential to supersede other ML algorithms with regards to the classification of helpful and unhelpful reviews. Hence, based on the obtained results, authors recommend the proposed model as an effective predictive tool that could be of immense help to the customers while making a purchasing decision about a particular product.

In the future, the scalability of the present work can be further enhanced and investigated on various heterogeneous datasets using ensemble learning models that might produce more robust and reliable outcomes from the perspective of commercial usability of the developed software tool.

Table 5. Comparative analysis of the best earlier models

Model	Dataset	Features	ML Algo.	Precision	Recall	Accuracy	F1 score	AUC
	Amazon	Review	CNN	97.00%	100%	95.27%	96.00	0.993
(Krishnamoorthy, 2015)	Amazon	Review	RandF	X	X	81.33%	87.21	X
(Malik et al. 2017)	Amazon	Review, Product	DNN	X	X	84.54%	92.1	X
(Bilal et al. 2021)	yelp	Review, Business, Reviewer	B-GBT	69.22%	68.25%	70.36%	68.73%	0.773

DATA AVAILABILITY

On request, Data shall be made available on reasonable request.

CONFLICTS OF INTEREST

There are no conflicts of Interest among authors.

ACKNOWLEDGMENT

No funding has been received for this research work.

REFERENCES

Alam, A., Singh, L. (2021). Distance-based confidence generation and aggregation of classifier for unstructured road detection, *Journal of King Saud University - Computer and Information Sciences*, 1319–1578. 10.1016/j.jksuci.2021.09.020

Alsmadi, A., AlZu'bi, S., Hawashin, B., Al-Ayyoub, M., & Jararweh, Y. (2020). Employing Deep Learning Methods for Predicting Helpful Reviews. *11th International Conference on Information and Communication Systems (ICICS)*, IEEE. doi:10.1109/ICICS49469.2020.239504

Bilal, M., Marjani, M., Hashem, I. A. T., Malik, N., Lali, M. I. U., & Gani, A. (2021). Profiling reviewers' social network strength and predicting the "Helpfulness" of online customer reviews. *Electronic Commerce Research and Applications*, *45*, 101026. doi:10.1016/j.elerap.2020.101026

Cao, J., Lin, Y., Huang, C. Y. Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. EMNLP-CoNLL (pp. 334–342).

Cao, Q., Duban, 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. doi:10.1016/j.dss.2010.11.009

Chau, V., Duong, D., Nguye, D., & Cao, T. (2018). From Helpfulness Prediction to Helpful Review Retrieval for Online Product Reviews, Association for Computing Machinery. ACM.

Chen, C. C., & Tseng, Y.-D., (2011). Quality evaluation of product reviews using an information quality framework, Decision Support Systems, *50*(4), 755e768.

Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *JMR, Journal of Marketing Research*, *43*(3), 345–354. doi:10.1509/jmkr.43.3.345

Chua, A. Y., & Banerjee, S. (2014). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*.

Dafni, R. J. (2022). VijayaKumar, K., Singh, L, Sharma, SK. (2022). Computer-aided diagnosis for breast cancer detection and classification using optimal region growing segmentation with MobileNet model. *Concurrent Engineering*, *30*(2), 181–189. doi:10.1177/1063293X221080518

Deepkiran, S. L., Pandey, M., & Lakra S. (2022, May). Automated Disease Detection in Plant Images using Convolution Neural Network, *International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, (pp. 487-492), doi:10.1109/CISES54857.2022.9844338

Filieri, R., Raguseo, E., & Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior*, *88*, 134–142. doi:10.1016/j.chb.2018.05.042

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*(3), 291e313.

Gang, R., & Hong T. (2018). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing and Management*, 0306-4573.

Ghose, A., & Ipeirotis, P. G., (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Transactions on Knowledge and Data Engineering, 23*(10), 1498e1512. IEEE .

Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, *52*, 101891. doi:10.1016/j.jretconser.2019.101891

Hu, N., Liu, L., & Zhang, J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology Management*, *9*(3), 201–214. doi:10.1007/s10799-008-0041-2

Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). Study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, *48*, 17–27. doi:10.1016/j.chb.2015.01.010

Huang, A. H., & Yen, D. C. (2013). Predicting the helpfulness of online reviews – A replication. *International Journal of Human-Computer Interaction*, *29*(2), 129–138. doi:10.1080/10447318.2012.694791

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, In: *Proceedings of the Eighth International Conference on Weblogs and Social Media*. AAAI . doi:10.1609/icwsm.v8i1.14550

Krishnamoorthy, S., (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications, 42*(7), 3751e3759.

Kumar, R., Singh, L., Tiwari, R. (2021). Path Planning for the Autonomous Robots Using Modified Grey Wolf Optimization Approach. *Journal of Intelligent and Fuzzy Systems, 40* (2021).

Lee, S., & Choeh, J. Y., (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications, 41*(6), 3041e3046.

Lee, S., & Choeh, J. Y. (2018). The interactive impact of online word-of-mouth and review helpfulness on box office revenue, Journel Management Decision. M.S.I. Malik, A. Hussain,"Helpfulness of product reviews as a function of discrete positive andnegative emotions. *Computers in Human Behavior*, *73*, 290–302.

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*(4), 101–136. doi:10.2753/JEC1086-4415170404

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, *56*(2), 172–184. doi:10.1016/j.im.2018.04.007

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*, 140–151. doi:10.1016/j.tourman.2014.09.020

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *Management Information Systems Quarterly*, *34*(1), 185–200. doi:10.2307/20721420

Murphy, R. (2020). Local Consumer Review Survey 2018. *Bright Local Research.* https://www.brightlocal.com/research/local-consumer-review-survey/(2020)

Otterbacher, J. (2009). Helpfulness in online communities: a measure of message quality. In: *Proceedings of the 27th SIGCHI Conference on Human Factors in Computing Systems* (pp. 955–964). ACM .

Park, Y. J. (2018). Predicting the Helpfulness of Online Customer Reviews across Different Product Types. *Computers in Human Behavior*, *88*, 132–148.

Quaschning, S., Pandelaere, M., & Vermeir, I., (2015). When consistency matters: The effect of valence consistency on review helpfulness. *Journal of Computermediated Communication, 20*(2), 136e152.

Ramanan, M., Singh, L., Kumar, A. S., Suresh, A., Sampathkumar, A., Jain, V., & Bacanin, N. (2022). Secure blockchain enabled Cyber- Physical health systems using ensemble convolution neural network classification. *Computers & Electrical Engineering*, *101*, 108058. doi:10.1016/j.compeleceng.2022.108058

S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M., (2006). Automatically assessing review helpfulness, In: *Proceedings of the Conference on empirical methods in natural language processing*. Association for Computational Linguistics.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*, 346–355. doi:10.1016/j.jbusres.2016.08.008

Singh, L., Alam, A., Kumar, K. V., Kumar, D., Kumar, P., & Jaffery, Z. A. (2021). Design of thermal imaging-based health condition monitoring and early fault detection technique for porcelain insulators using Machine learning. *Environmental Technology, and Innovation*, *24*(21), 102000. doi:10.1016/j.eti.2021.102000

Srinivas, K., Singh, L., Chavva, S. R., Dappuri, B., Chandrasekaran, S., & Qamar, S. (2022). Multi-modal cyber security-based object detection by classification using deep learning and background suppression techniques. *Computers & Electrical Engineering*, *103*, 108333. doi:10.1016/j.compeleceng.2022.108333

Wan, Y. (2015). The Matthew effect in social commerce. *Electronic Markets*, *25*(4), 313–324. doi:10.1007/s12525-015-0186-x

Wang, X., Tanga, L., & Kimb, E. (2018). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*.

Zhang, Z., & Varadarajan, B. (2006). Utility scoring of product reviews, In: *Proceedings of the 15th ACM international conference on information and knowledge management CIKM'06* (pp. 51–57). ACM.

Zhang, Z., Wei, Q., & Chen, G. (2014). *Estimating online review helpfulness with probabilistic distribution and confidence, Foundations and Applications of Intelligent Systems*. Springer.

Surya Prakash Sharma has 19 Years' experience in the field of academic and research. He has completed his B.Tech. in Computer Science & Engineering from Kanpur University (India) in 2003 and M.Tech in Computer Science & Engineering from M.D. University Rohtak (India) in 2011, respectively. Currently he is pursuing his Ph.D. degree from Dr. A.P.J. Abdul Kalam Technical University Lucknow (India). His research area are in machine learning, deep learning and opinion mining.

Laxman Singh obtained his B. Tech in Electronics and Communication Engineering from C.R. State (Govt.) College of Engineering, Murthal, Sonipat (Haryana) and M.Tech in Instrumentation and Control from M.D. University, Haryana, India in 2004 and 2009 respectively. He received his PhD degree from Jamia Millia Islamia (a central Govt. of India University) in 2016. Presently he is working as Associate Professor in the Department of Electronics & Communication Engineering at Noida Institute of Engineering & Technology (NIET), Greater Noida. He has total teaching experience of more than seventeen years. Dr. Laxman Singh has published about 70 research articles in the field of image processing, AI, machine learning, and robotic path planning in various refereed international/national journals as well as in international conferences of repute. He has authored two books; one on "Practical machine learning with Python", and another on "Electronic Circuits and applications", published by GBS Publisher, New Delhi. His current research interests are in the areas of Wavelet analysis, Artificial Intelligence, Image processing, Optimization techniques, and robotic path planning.

Rajdev Tiwari is a committed academician with around 20 years of experience at premier engineering colleges of Uttar Pradesh like KNIT Sultanpur, IPEC Ghaziabad, ABESIT Ghaziabad, NIET Greater Noida etc. Currently he is working with GNIOT Greater Noida as Professor and Head Computer Science and Engineering department. He is basically M.Sc in Electronics, MCA and have done PhD in Computer Science. He has also done PGDASDD from CDAC Noida and qualified UGC NET in year 2012. He has been visiting faculty at AMITY University, Noida and at IETE, New Delhi for PhD and ALCCS programs respectively for past many years. He is actively associated with various professional bodies like IEEE, CSI, ISTE etc. He has published more than 30 research papers in various journal of international repute. He has attended and chaired various international conferences. He has delivered keynote speeches at various TEQIP-III FDPs and judged many project exhibitions as technical judge. He has guided around 10 M.Tech dissertations and two PhD. He is guiding 5 PhD scholars from Dr APJAKTU, at present. He has authored two books; one on soft computing published by Acme Learning, and another on Algorithms published by Pearson. He is on the panel of various International Journals as Editor/Reviewer. He has received various grants for Research Project, Conferences & FDPs from reputed organizations like Dr APJ AKTU, Lucknow & AICTE, New Delhi.