International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 14 Issue 6, June-2025

Sentiment Analysis of Product Reviews using Machine Learning

Mahasweta Saha[1], Jishu Chakraborty[2], Samarendra Das[3], Sonali Bhowmik[4], Prof. (Dr.) Sumit Nandi[5]

[1,2,3]Student, Department of Computer Science and Engineering, JIS University, West Bengal, India
[4]Assistant Professor, Department of Computer Science and Engineering, JIS University, West Bengal, India
[5]Harishchandrapur College, Pipla, Malda, West Bengal

ABSTRACT

Digital reviews have become crucial in boosting worldwide communication among customers and significantly influence purchasing decisions. Platforms like Amazon and Flipkart allow customers to share their experiences, providing authentic insights into product performance for potential consumers. To extract meaningful insights from a large number of reviews, it is crucial to categorize them into positive and negative sentiments. Sentiment Analysis, a approach for extracting subjective information from text, is utilized in this process. In this study, we analyse over 9374 reviews and classify them into positive and negative sentiments using Sentiment Analysis. We applied various classification models, including Naïve Bayes, Logistic regression, Random Forest and K-Nearest Neighbors (KNN), to perform this classification.

Keywords: Big data, Sentiment Analysis, Opinion mining, Naïve Bayes, Logistic regression, Random Forest, KNN, Machine learning.

INTRODUCTION

With the growing demand for headphones, the market is expanding rapidly. This surge in the headphone industry calls for a comprehensive review of different brands and models. Numerous brands dominate the market, with names like Oneplus, Boat, Realme, Sony recognized worldwide.

E-commerce acts a pivotal role in boosting headphone sales and impacting consumer buying habits. Reviews on platforms like Flipkart.com help guide consumers in making informed decisions. Shoppers can give a numerical rating from 1 to 5 or write detailed reviews about their purchases. Given the huge number of products from various brands, relevant and accurate reviews are essential.

The sheer volume of reviews can be overwhelming, similar to handling big data. To help manage this, classifying reviews by sentiment—positive or negative—can provide clearer insights for potential buyers. This classification helps future customers evaluate feedback more constructively and make better decisions based on their needs.

In this research, we've extracted unstructured headphone review data from Flipkart.com. We filtered out the noisy data and pre-processed it to calculate the sentiment of the reviews using supervised learning and unsupervised learning. We then classified the reviews using several machine learning models, including Naïve Bayes, Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest.

By organizing reviews this way, potential buyers can quickly see the pros and cons of different headphones, helping them make more informed choices. This process also enhances user credibility, as it provides detailed and trustworthy feedback about headphone products.

RELATED WORK

Data analytics has revolutionized our ability to uncover implicit patterns in vast amounts of data. Big Data is characterized by Volume, Velocity, and Variety, but Veracity and Value are also crucial components[1]. The massive volume and rapid generation of data every day surpass the computational capabilities of numerous IT departments. E-commerce websites, for example, are filled with various reviews for various products, which can be analyzed to understand customer behaviour and make decisions. These reviews can be structured or unstructured, and by filtering out irrelevant data, we can extract beneficial business perceptions. Big Data has enabled businesses to prosper and enhance operations based on evidence rather than intuition. It aids in targeted social influencer marketing, customer segmentation,

identifying sales and marketing opportunities, fraud detection, risk quantification, planning and forecasting, and understanding consumer behavior [2].

Sentiment analysis involves identifying the sentiment of reviews and categorizing them as positive, negative, or neutral. This analysis can be conducted at the document, sentence, or phrase level [3]. Prior research has classified words and phrases with inherent positive or negative polarities [4]. While this prior classification is helpful, contextual polarity can significantly alter the meaning. By considering contextual polarity, ambiguity is reduced, and the intended sentiment is preserved [5][6].

Extensive studies have been performed on tweets and movie reviews to establish the groundwork for sentiment analysis and opinion mining. Sentiment classifiers have been developed to classify sentiments not only in English but also in other languages using data from platforms like Twitter [7]. For instance, smartphone product reviews have been assessed based on their positive and negative orientations [8]. Advanced sentiment analysis systems, such as those using Support Vector Machines (SVM), account for sarcasm, grammatical errors, and spam detection [9]. Enhanced Naïve Bayes models have improved sentiment analysis by incorporating methods like effective negation handling, word n-grams, and feature selection [10].

Sentiment analysis extends beyond English to other languages. For example, sentiment analysis of Chinese text has been conducted using multiple classifiers, including Centroid classifier, K-nearest neighbor, Window classifier, Naïve Bayes, and SVM, with SVM often outperforming others [11]. In travel reviews, SVM and character-based N-gram models have shown better execution than Naïve Bayes [12]. In most cases, SVM typically provides the best results compared to other classification models.

Overall, sentiment analysis is not limited to critiques or Twitter data but is also relevant to stock markets [13][14], news articles [15] and more. This broad applicability underscores its importance in extracting meaningful insights from large datasets across various domains.

METHODS

The suggested method of the work is done by following three various segments.

Dataset and its Attributes

The initial step involves data collection and preprocessing. A huge amountof online reviews is gathered from the e-commerce website Flipkart.com. The data set consists of more than 9374 reviews of earphones and headphones. It includes nine attributes as mentioned in Table 1.

Table 1: Features of the Dataset

Feature	Description
Product_id	Identity number of the Product
Product_name	Model name of the product along with manufacturing brand
Rating	User rating between 1 to 5
Summary	Review title based on user's rating
Review	User reviews provided for each product
Location	User locations
Date	On which date the product was bought
Upvotes	Total Number of upvotes received
Downvotes	Total Number of downvotes received

Approach

The proposed framework adheres to a systematic approach outlined in Fig. 1. First, we collected experimental data from the e-commerce website Flipkart.com. The data sets are in Comma Separated Values (CSV) format and are available as supplements.

Next, we pre-processed the data to clean it up by removingstop words, punctuation marks, whitespaces, digits, and special symbols using ‘nltk’ package. After that, we performed feature selection to derive the most key sights from the data set. Out of nine features, we focused on three: Product Name, Rating, and Review.

Next, we determined the sentiment polarity of the reviews. Then, we tagged each review with 'Positive' or 'Negative' labels to facilitate supervised learning. Finally, we trained and tested the classified data using Naïve Bayes, Logistic regression, Random Forest and K-Nearest Neighbors (KNN).

Sentiment Analysis

The process for analyzing sentiment in product reviews are shown in figure 1. as follows

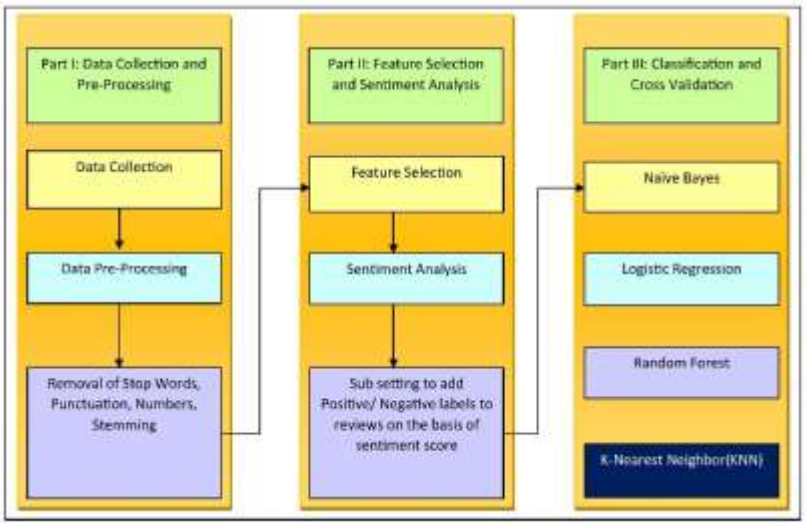

Fig. 1 Proposed Method

Here, we focused solely on positive and negative sentiment orientations to classify the reviews. Figure 2 shows the ratio of positive and negative reviews in the dataset, visualized with a bar graph.

Classifications

Classification involves leveling reviews based on their sentiment into two classes: Positive and Negative. Using the Vader, we calculated sentiment scores for each positive and negative review. These scores are then included in the reviews dataset. The overall sentiment polarity is calculated using the formula shown in Fig. 3, as given by Eq. 1:

Sentiment Score = Positive Score – Negative Score

This formula helps determine the overall sentiment of each review.

After calculating the polarity of each review, we add a new feature based on their sentiment. The reviews are marked as 'Positive' and 'Negative' Based on their polarity score. Positive scores are scaling and Negative scores are scaling from -1 to -10. These subsets are then combined in a CSV file along with their corresponding reviews.

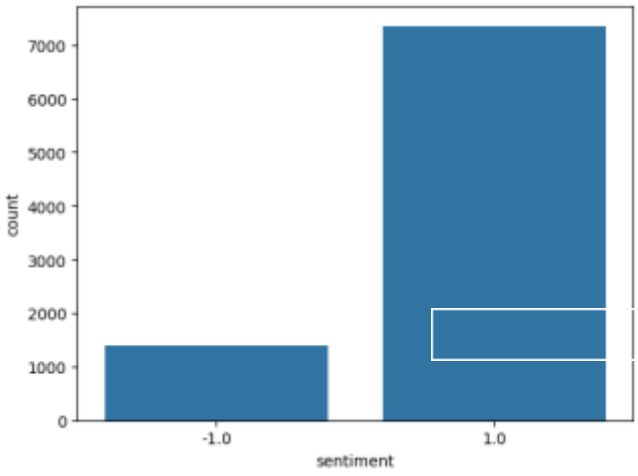

Fig. 2. Total Positive and Negative Reviews

rating	review	Sentiment_score	sentiment
5	1-more flexible2-bass is very high3-sound clarity i	0.268181818	Positive
5	Super sound and good looking I like that prize	0.477777778	Positive
5	Very much satisfied with the device at this price p	0.536111111	Positive
5	Nice headphone, bass was very good and sound is	0.346	Positive
5	Sound quality super battery backup super quality s	0.346666667	Positive
5	Wowwww it's amezing bluetooth nice look, nice p	0.46	Positive
4	Awesome colour! Amazing experience .. but only	0.424285714	Positive
5	For the first time, I am posting a review, just beca	0.275714286	Positive
4	First of all the Delivery boy is a good guy. Nice rec	0.441666667	Positive
1	This headphone is good but not that much as i exp	0.308571429	Negative
5	It's a very good product ... I'm so happy with this I	0.432222222	Positive
4	Nice build quality.. bass is best it self.. vibrate at c	0.53	Positive
1	After using for about 8 days, right side of the earp	0.142857143	Negative
1	Mic is not proper recipient persons can't hear pro	0.07	Negative
5	Wow flip cart super fast service like express spee	0.393518519	Positive
5	Very great product at this price I will assure that t	0.58	Positive
4	It's very good and its bass is very high This bluetoc	0.35155	Positive
5	Good product. Never used bluetooth headset bef	0.551851852	Positive
5	This review is after 1 month of use .Pro:- 1.Batter	0.272571429	Positive
5	I like It Boat HP Sound Quality is bestm best And B	0.426623377	Positive

Fig. 3: Sentiment score of the review

Classifiers
The classification models are used to classify reviews are : Naïve Bayes, Logistic regression, Random Forest and K-Nearest Neighbors (KNN).

Naïve Bayes
Naïve Bayes is a classification model based on Bayes' Theorem, which implies independence among factors. This model is easy to build and particularly beneficial for handling extensive datasets. It works on Bayes' theorem of probability to predict the class of unknown datasets. Essentially, it is a statistical classifier that assigns input feature vectors to output class labels.

For a set of training data D, each row is represented by an n-dimensional feature vector, $X = (x_1, x_2, \dots, x_n)$. There are K classes, $K_1, K_2, \dots, K_m$, in the output class label. For every tuple X, the classifier predicts that X belongs to K_i if and only if:

$$P(K_i|X) > P(K_j|X)$$

where $i, j \in [1, m]$ and $i \neq j$. The probability $P(K_i|X)$ is calculated as:

$$P(K_i|X) = \prod^{n}_{a=1} P(x_a|K_i)$$

This formula means that the probability of X belonging to class K_i is the product of the probabilities of each feature x_a given K_i.

Logistic Regression
Logistic regression is utilized to forecast the probability of an outcome with only two possible values (a dichotomy). It makes predictions using one or more predictors, which can be either numerical or categorical. Linear regression is ineffective for predicting binary outcomes for two key reasons:

1. Linear regression can predict values outside the appropriate range, such as probabilities exceeding the 0 to 1 range.
2. The residuals would not necessarily spread around the expected axis since binary outcomes can only have one of two possible values for each observation.

In contrast, logistic regression creates a logistic curve with values exceeding from 0 to 1. Instead of directly using probabilities, logistic regression uses the natural logarithm of the target variable's "odds" to construct the curve. Additionally, the predictors in logistic regression do not need to be normally distributed or have the same variance across categories to be effective.

Random Forest
Random Forest is an effective procedure that combines multiple decision trees to improve classification performance. When using single tree classifiers like decision trees, factors like outlier data or noisy data can detrimentally affect the classifier's accuracy. Random Forest addresses these issues by introducing randomness, making it highly resistant to noise and outliers.

This classifier model generates two types of randomness: data randomness and function randomness. It also has several hyperparameters to tune, such as:

- The number of trees to build in the forest.
- The maximum number of features to select at random.
- The maximum height of each tree.

Because it uses the principles of bootstrapping and bagging, Random Forest is considered a reliable and accurate classifier.

K-Nearest Neighbors (KNN)
For this approach, we selected KNN (K-Nearest Neighbors) as the classifier. Given that sentiment analysis is a binary classification task and involves large datasets, KNN is well-suited for our needs. We used a manually leveled training set to train the classifier. In the training set, there is an X:Y relationship where X represents the score of the review, and Y indicates whether the review is positive or negative. The score of the review related to a feature within the review is provided as input to the KNN classifier.

RESULTS AND DISCUSSIONS

In this study, the input consists of customer reviews, and the output is the sentiment column, derived from the rating column. The reviews undergo data pre-processing before feature vectors are created using vectorization techniques like Count Vectorizer. These feature vectors are then used as input for machine learning models such as Naïve Bayes, Logistic Regression, Random Forest, K-Nearest Neighbors (KNN). We use 80% of the dataset to train and 20% to test the models. After training, the machine learning models predict the sentiment of the reviews.

Performance Evaluation
After testing and training the dataset, we calculate the predictive accuracy to determine the best model. Among them, the Random Forest has highest predictive accuracy among all of them.

Table 1: Model Accuracy

Model Name	Accuracy
Naïve Bayes	96.32%
Random Forest	97.37%
Logistic Regression	96.17%
K-Nearest Neighbors	96.42%

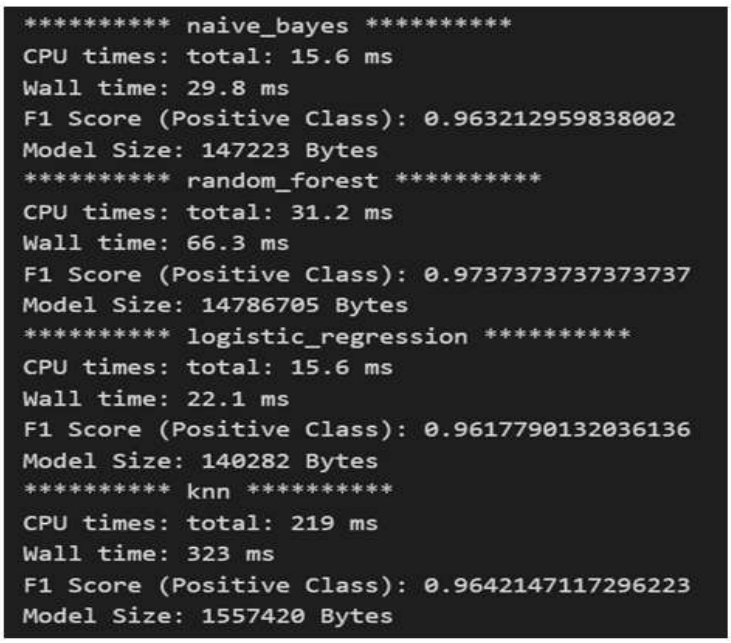

Fig. 3. Comparison between Classifier models

Equation/Formula

1. Accuracy (ACC):

ACC = {TP + TN} / {TP + TN + FP + FN}

Where:

- TP is the number of true positives (correct identifications).
- TN is the number of true negatives (Accurate non-positives).
- FP is the number of false positives (Incorrect identifications).
- FN is the number of false negatives (incorrect rejections).

2. Sensitivity (Recall):

Sensitivity = {TP} / {TP + FN}

Sensitivity is also known as recall or true positive rate. It evaluates the ability of the model to accurately detect positive instances.

3. F1 Score:

F1 Score = {2 * Precision * Recall} / {Precision + Recall}

Where:

Precision = {TP} / {TP + FP} is the precision or positive predictive value.

The F1 score is the Interquartile mean of precision and recall and provides a stability between the two metrics.

4. Specificity:

Specificity = {TN} / {TN + FP}

Specificity evaluates the ability of the model to Accurately detect negative instances.

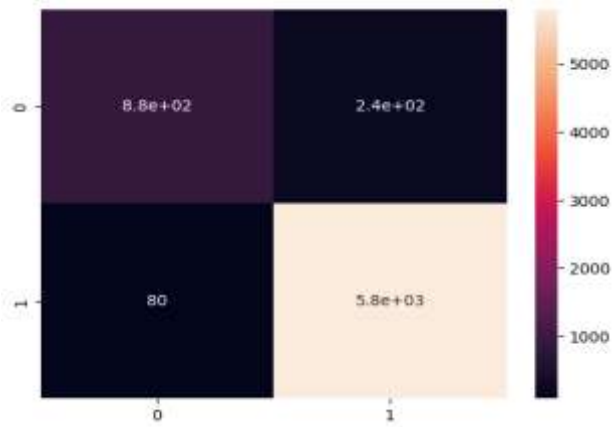

Fig. 4. Confusion Matrix of Naïve Bayes

Fig. 5. Output of a review

CONCLUSION

Online reviews play a pivotal role for fostering trust and persuading customers in the field of digital market. With so many reviews out there, it's essential to manage this information and effectively showcase authentic reviews to consumers. Our research aims to do just that by performing sentiment analysis on earphone reviews and sorting them into positive and negative sentiments.

After sorting the data as positive and negative reviews, we used four classification models to sort the reviews. We used Naïve Bayes, Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest for this task. Among these, Random Forest model showed the best predictive accuracy. The accuracy results were cross-validated, with the highest accuracy achieved being 97.44% for the Random Forest model.

Looking ahead, this work can be expanded to execute multi-class classification of reviews. This would give more detailed insights into the nature of each review, helping consumers make better decisions about products. Additionally, it could predict a product's rating based on the review content, offering users more reliable ratings. Sometimes, the rating given to a product and the sentiment of the review don't align perfectly, and this approach could help bridge that gap. This proposed extension would be highly beneficial for the e-commerce industry, enhancing user satisfaction and trust by providing more accurate and detailed feedback.

REFERENCES

[1]. S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," Journal of BusinessResearch, Vol. 69, no. 2, pp. 897–904, 2016.
[2]. P. Russom et al., "Big data analytics," TDWI best practices report, fourth quarter, pp. 1–35, 2011.
[3]. V. Umarani, A. Julian, and J. Deepa. "Sentiment analysis using various machine learning and deep learning Techniques." Journal of the Nigerian Society of Physical Sciences (2021): 385-394.
[4]. V. Hatzivassiloglou and K. R. McKeown, "Predicting the se-mantic orientation of adjectives," in Proceedings of the eightconference on European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 1997,pp. 174–181.
[5]. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005, pp. 347–354.
[6]. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004, p. 271.
[7]. A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining." in LREc, vol. 10, no. 2010, 2010.
[8]. M. Wahyudi and D. A. Kristiyanti, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." Journal of Theoretical & Applied Information Technology, vol. 91, no. 1, 2016.
[9]. D. N. Devi, C. K. Kumar, and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in Advanced Computing (IACC), 2016 IEEE 6th International Conference on. IEEE, 2016, pp. 3–8.
[10]. V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive bayes model," in International Conference on Intelligent Data Engineering and Automated Learning. Springer, 2013, pp. 194–201.
[11]. S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents," Expert Systems with applications, vol. 34, no. 4, pp. 2622–2629, 2008.
[12]. Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine
earning approaches," Expert Systems with Applications, vol. 36, no. 3, pp. 6527–6535, 2009.
[13]. L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their for the sentiment classification of stock market news," Knowledge-Based Systems, vol. 41, pp. 89–97, 2013.
[14]. M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," Decision Support Systems, vol. 55, no. 3, pp. 685–697, 2013.