JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Sentimental Analysis Of Amazon Electronic Product Reviews Using Naive Bayes On Alexa

**[1.] Dr. N. Rahul Pal, [2.] Mrs. S. Sadhiya**

[1.] Assistant Professor, [2.] Student

[1.] Adikavi Nannaya University, Rajamahendravaram, India

[2.] Adikavi Nannaya University, Rajamahendravaram, India

## ABSTRACT

In the digital age, sentiment analysis has emerged as a crucial technique for comprehending consumer opinions and feedback. This research article employs the Naive Bayes technique to apply sentiment analysis to Amazon electronic product reviews, specifically for Amazon's Alexa devices. By dividing product reviews into three categories—positive, negative, and neutral—the study hopes to gain a better picture of how customers feel about Alexa's functionality, performance, and usability. Because of its ease of use, effectiveness, and capacity to manage huge datasets, the Naive Bayes classifier—a probabilistic machine learning algorithm—is employed. The study takes a methodical approach, first gathering a sizable dataset of Alexa product reviews from Amazon and then preparing the data. Naive Bayes is used to classify sentiment, and evaluation criteria including accuracy, precision,recall and F1-score, which are used to gauge the model's effectiveness. The results of this study advance the field of sentiment analysis in e-commerce and provide Amazon and other businesses with useful information to enhance their product offers based on customer input. In the end, the study shows that it is feasible to process and analyze customer evaluations in a useful, scalable way using machine learning techniques.

**Keywords:***Sentimental Analysis, Naïve Bayes Algorithm*.

## INTRODUCTION

Online customer reviews are now a crucial component of consumers' decision-making processes in today's digital environment. Platforms like Amazon have emerged as the preferred source for product purchases due to the expanding trend of e-commerce, and consumers frequently utilize reviews to evaluate the performance and quality

of products. Because of their popularity and inventiveness, electronic devices—especially smart home assistants like Amazon Alexa—have attracted a lot of attention among the vast array of products sold on Amazon. Manufacturers, retailers, and consumers themselves can all benefit from knowing the attitude expressed by customers in reviews of these items.

## Machine Learning:

This approach works on the machine learning technique. Firstly, the datasets are trained and predictive analysis is done. The next process is the extraction of words from the text is done. This text extraction can be done using different techniques such as Naïve Bayes,support vector machines, hidden Markov model, and conditional random fields like this machine learning techniques are used.

## Sentimental Analysis:

- Opinion mining, another name for sentiment analysis, is a branch of natural language processing (NLP) that focuses on examining textual material to ascertain its sentiment or emotional tone.
- By categorizing reviews as neutral, negative, or positive, this method gives organizations insightful information that helps them assess consumer satisfaction, pinpoint areas for development, and adjust their marketing tactics.
- Sentiment analysis can provide insights into how customers feel about particular features, functionality, usability, and general satisfaction with Amazon Alexa and other electronic items.

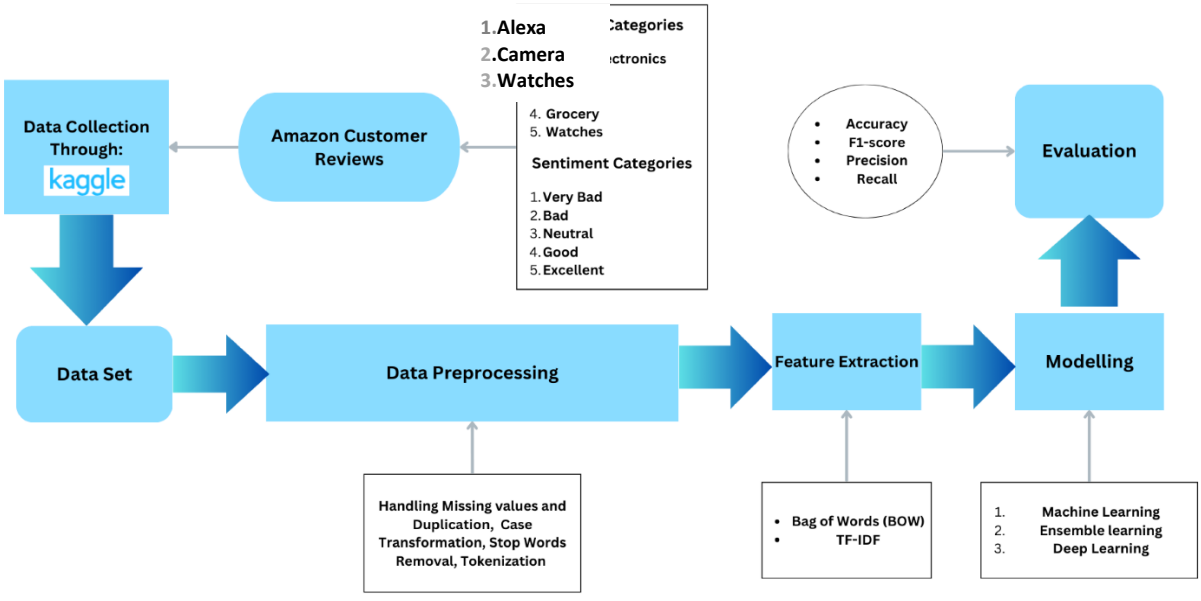


**Figure 1:** Proposed Framework

Sentiment analysis tasks have demonstrated the immense promise of machine learning methods, especially supervised learning models. Because of its ease of use, effectiveness, and efficiency in managing big datasets with comparatively low processing requirements, the Naive Bayes classifier is one of the most popular among them. It is the best option for examining customer evaluations on e-commerce sites like Amazon since it excels at text classification problems.

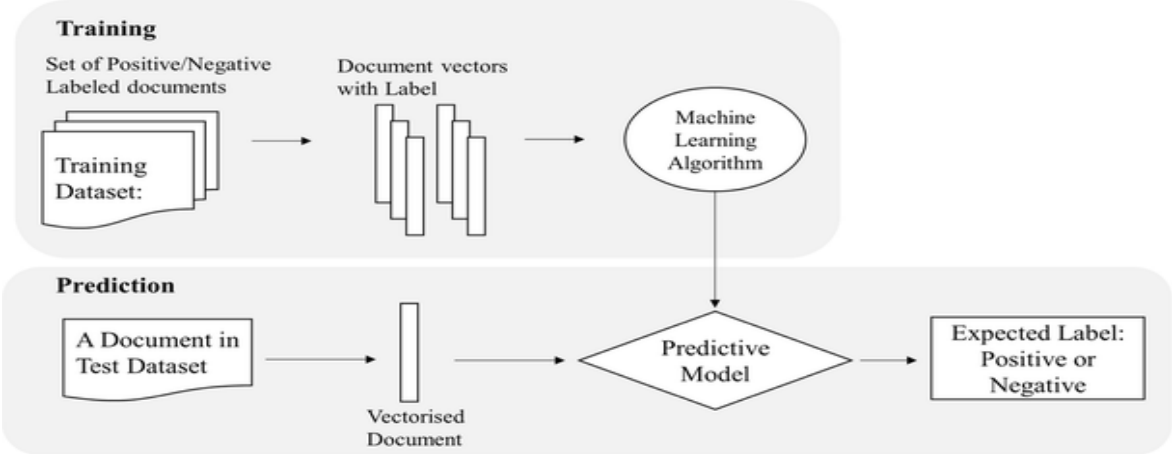


**Figure 1:**A simple example of a supervised machine learning method for sentiment classification.

## Naïve Bayes Algorithm:

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.
- It is mainly used in text classification that includes a high-dimensional training dataset.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

This study attempts to apply the Naive Bayes algorithm for sentiment analysis of Amazon electronic product reviews. In order to provide insights on the user experience and satisfaction with Amazon's Alexa products, the study analyses a sizable dataset of customer reviews and attempts to categorize them into good, negative, or neutral categories. Additionally, using the textual data from the reviews, the study will assess the Naive Bayes classifier's performance and determine how well it can predict consumer sentiment.

# II) LITERATURE REVIEW

## [1] Recent Research on Sentiment Analysis in Electronic Product Reviews

- **Hossain (2021)**: Focused on sentiment analysis of electronic product reviews using different machine learning algorithms, including Naïve Bayes. They found that Naïve Bayes, when combined with specific text preprocessing techniques, outperformed other classifiers like Decision Trees and SVM.
- **Mohammad (2020)**: Proposed a hybrid approach that combined Naïve Bayes with deep learning techniques for more accurate sentiment classification. They showed that hybrid models can overcome some of the limitations of Naïve Bayes, particularly when dealing with complex sentence structures and sarcasm.

## [2] Sentimental Analysis of Product reviews:

**Authors: Najma Sultan & Pintu Kumar & Monika Rani Patra & Sourabh Chandra and S.K. Safikul Alam(2019)**

This research paper aims at running sentimental analysis on customer reviews and labeling text as either "Positive", "Negative" or "Neutral". The paper discusses a theoretical approach to sentimental analysis and analyses different algorithms for the same with their corresponding accuracies. It also gives a brief history of different other approaches to sentimental analysis techniques. Training a dataset to classification algorithms like Naive Bayes classification algorithm, Linear Model algorithm, SVM algorithm and Decision tree is also done to compare accuracies. In the testing phase, the user review is mapped to the saved feature set. Feature extraction is done using the frequency of occurrence.

# III) Existing System Overview:

## 3.1. Traditional Sentiment Analysis Systems

Conventional sentiment analysis methods use more straightforward machine learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), and Logistic Regression, to extract sentiments from product reviews. The following stages are typically taken by these systems:

- **Data Collection:** Using web scraping tools or APIs, current systems collect reviews from websites such as Amazon. A variety of technological gadgets, including Alexa devices, may be the subject of the reviews.

- **Text Preprocessing:** To get the text ready for analysis, data cleaning and preprocessing methods like lemmatization, stemming, tokenization, and stopword removal are used.
- **Feature Extraction:** To transform text input into numerical vectors that machine learning models can analyze, these systems employ techniques like Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), or word embeddings.
- **Sentiment Classification:** The processed text is subjected to machine learning classifiers, such as Naive Bayes, which classify reviews into predetermined sentiment classes, such as positive, neutral, and negative. Naive Bayes classifies sentiment using probabilistic computations and the assumption that words (features) are independent of one another.

# IV) Proposed System Overview

By combining Alexa with Naive Bayes sentiment analysis, the suggested system seeks to enhance the current ones and provide a more engaging, user-friendly experience when examining Amazon product reviews. The following elements make up the suggested system:

## 4.1. Proposed System Components

### 1. Enhanced Data Collection

- **Attention to Alexa Product Reviews:** The suggested system will extract a more focused set of reviews with an emphasis on Alexa devices. Reviews for different Alexa models (Echo, Dot, Show, etc.) will be included in this focused collection.
- **Managing Review Metadata**: To improve sentiment analysis and weed out unhelpful or low-quality reviews, the suggested method would consider extra metadata such verified purchase status, review date, and scores.

### 2. Advanced Text Preprocessing

- **Context-Aware Preprocessing**: Unlike traditional systems, the proposed system will introduce more sophisticated preprocessing methods to handle domain-specific language, sarcasm, and mixed sentiments. For example:
- Use contextualized word embeddings like BERT (Bidirectional Encoder Representations from Transformers) for more accurate understanding of words in context.

**3. User-Friendly Interface**

- **Visual Sentiment Dashboard**: The proposed system will include a user interface that displays sentiment analysis results in an intuitive manner, with sentiment trends over time, word clouds, and key sentiment drivers for Alexa reviews.

**4. Alexa Integration for Real-Time Feedback**

- **Custom Alexa Skill**: The proposed system will involve creating a **custom Alexa Skill** that integrates with the sentiment analysis model. Users can ask Alexa for the sentiment of product reviews in real-time.
- **Voice Output**: Alexa will provide a voice-based summary of the sentiment, e.g., "Most reviews are positive, praising the camera quality and display, but some users have complained about battery life."

# V) METHODOLOGY

The process used to perform sentiment analysis on Amazon products Reviewing Alexa usually entails a number of important phases, from gathering data to assessing the model. Here is a thorough rundown of the procedure:

## i)Data Collection:

Kaggle is the source of the dataset. The general dataset analysis methods are used to analyse the dataset.

**Table 1:** Details of the dataset

| | |
|---|---|
| Dataset consists | 2000 reviews |
| Rows | 2000 |
| Columns | 5 |
| Attributes | Rating, dates, variation, verified reviews, feedback |

## Dataset Overview:

**Table 2:** Features of a dataset

| | |
|---|---|
| Review text | Contains the text of review. |
| Overall | Rating given by the customer (1-5 scale). |

| | |
|---|---|
| Sentimental labelling | Ratings 1-2 as negative<br>Ratings 4-5 as positive<br>Excluding rating 3(Neutral) |

## ii) Data Preprocessing:

The reviews will undergo several preprocessing steps to ensure the quality of input data for the Naïve Bayes model:

- **Tokenization**: Splitting the text into words or tokens. These tokens can be individual words, punctuation marks, or even sub-word units like characters or sub words.
- **Stop-word Removal**: Eliminating common words (e.g., "the," "and," "is") that do not contribute to sentiment.
- **Stemming and Lemmatization**: Reducing words to their base forms (e.g., "running" to "run").
- **Lowercasing**: Converting all upper case letters text into their corresponding lowercase letters to maintain uniformity.
- **Noise Removal**: Removing special characters, numbers, and unnecessary punctuation marks.

## iii) Naïve Bayes Classification:

The Naïve Bayes classifier determines the likelihood that a review falls into one of three sentiment classes—positive, negative, or neutral—and then chooses the class with the highest likelihood.

1. **Training**: The Naïve Bayes model will be trained on a labeled dataset, where each review is associated with a known sentiment label.
2. **Prediction**: The model will then classify new, unseen reviews into one of the sentiment classes (positive, negative, neutral) based on the probabilities calculated during training.
3. The following is how the naive bayes classifier works: consider set of training data. D, where every tuple is represented thru n-dimensional trait vector. A = a1, a2, a3, ... a. indicate n measurement performed upon tuple using n attributes or else features. suppose there is b class in total, D1, D2, D3, ... If and only if, the classifier will predict that tuple A belongs to $C_1$: P(DIA) > P(D/A), where k. / [1, b] and k is not equal to 1.

The classification will use a multinomial Naïve Bayes approach, which is well-suited for text classification tasks.

## iv) Model Training:

- **Split Data:** Divide the dataset into training, validation, and test sets to evaluate model performance**.**
- **Training:** Train the selected model on the training dataset using labelled sentiments (positive, negative, neutral).

### v) Model Evaluation:

To evaluate the performance of the Naïve Bayes classifier, the following metrics will be used:

- **Precision, Recall, F1-Score**: These metrics will provide deeper insights into the model's ability to classify each sentiment class accurately.
- **Confusion Matrix**: This will help visualize the true positives, false positives, true negatives, and false negatives for each class.
- **Performance Metrics:**Evaluate model accuracy using metrics such as accuracy, precision, recall, F1-score, and confusion matrix**.**
- **Validation:**Use the validation set to fine-tune hyperparameters and improve model performance.

### vi) Sentiment Classification:

- **Prediction:** Apply the trained model to classify the sentiments of unseen restaurant reviews.
- **Aspect-Based Analysis:** Optionally, conduct aspect-based sentiment analysis to evaluate sentiments related to specific features (e.g., quality, service, ambiance).

### vii) Visualization and Reporting:

- **Data Visualization:** Use tools like Matplotlib or Seaborn to create visual representations of sentiment trends and insights**.**
- **Reporting:** Summarize findings in reports or dashboards to provide actionable insights for product management

## VI) EXPERIMENTAL RESULTS

```
#importing the required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
import re
from wordcloud import WordCloud
from nltk.corpus import stopwords
```

**Figure 3:**Importing the libraries.

```
#importing the required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
import re
from wordcloud import WordCloud
from nltk.corpus import stopwords

#load the dataset
df = pd.read_csv('/kaggle/input/amazon-alexa-reviews/amazon_alexa.tsv', sep='\t')
df.head()
```

**Figure4:**Loading dataset of product.

```
df.shape
df.info()
#check for null values
df.isna().sum()
#drop the missing value row
df.dropna(inplace=True)
df.shape
#create a new column 'length' for length of the reviews
df['lenght'] = df['verified reviews'].apply(len)
df.head()
```

**Figure 5:** Data preprocessing.

| | rating | date | variation | verified_reviews | feedback | length |
|---|---|---|---|---|---|---|
| 0 | 5 | 31-Jul-18 | Charcoal Fabric | Love my Echo! | 1 | 13 |
| 1 | 5 | 31-Jul-18 | Charcoal Fabric | Loved it! | 1 | 9 |
| 2 | 4 | 31-Jul-18 | Walnut Finish | Sometimes while playing a game, you can answer... | 1 | 195 |
| 3 | 5 | 31-Jul-18 | Charcoal Fabric | I have had a lot of fun with this thing. My 4 ... | 1 | 172 |
| 4 | 5 | 31-Jul-18 | Charcoal Fabric | Music | 1 | 5 |

**Figure 6:**Output after dataset cleaning.

```
df['rating'].value_counts()
#plotting the counts of 'rating' values
sns.countplot(x='rating', data=df)
```

```
<Axes: xlabel='rating', ylabel='count'>
```

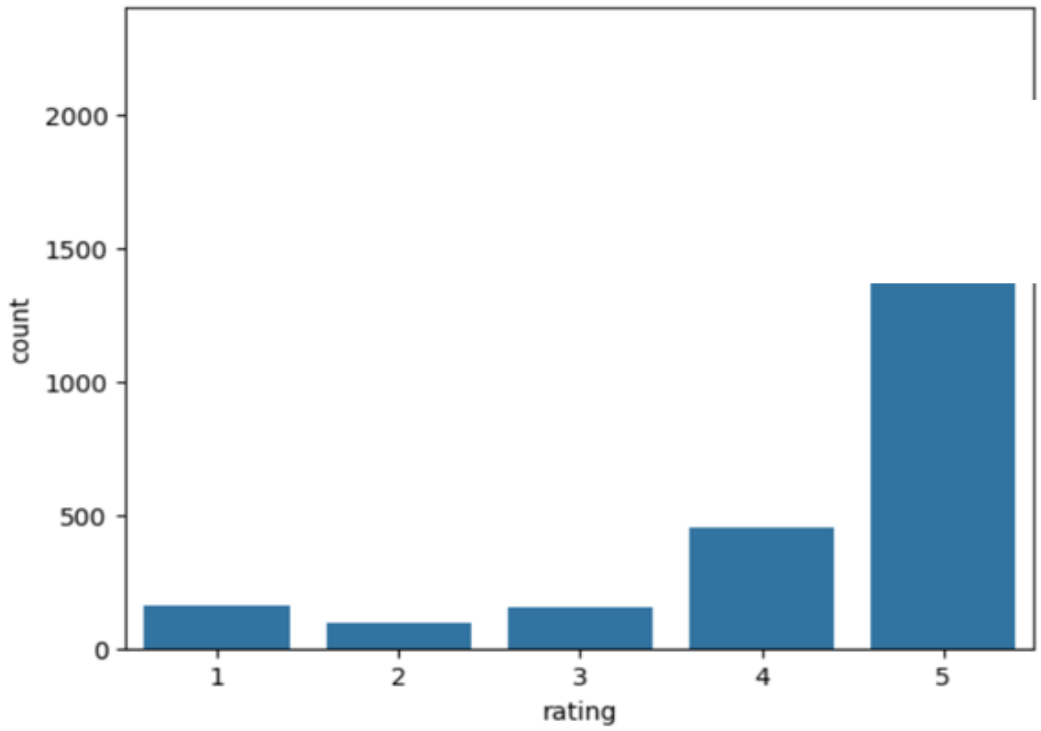


**Figure 6:**Analysing the "rating" column

```
wc = WordCloud(background_color='white', max_words=50)
plt.figure(figsize=(10,10))
plt.imshow(wc.generate(review_string))
plt.axis('off')
plt.show()
```

**Figure 7:** Generating Word cloud

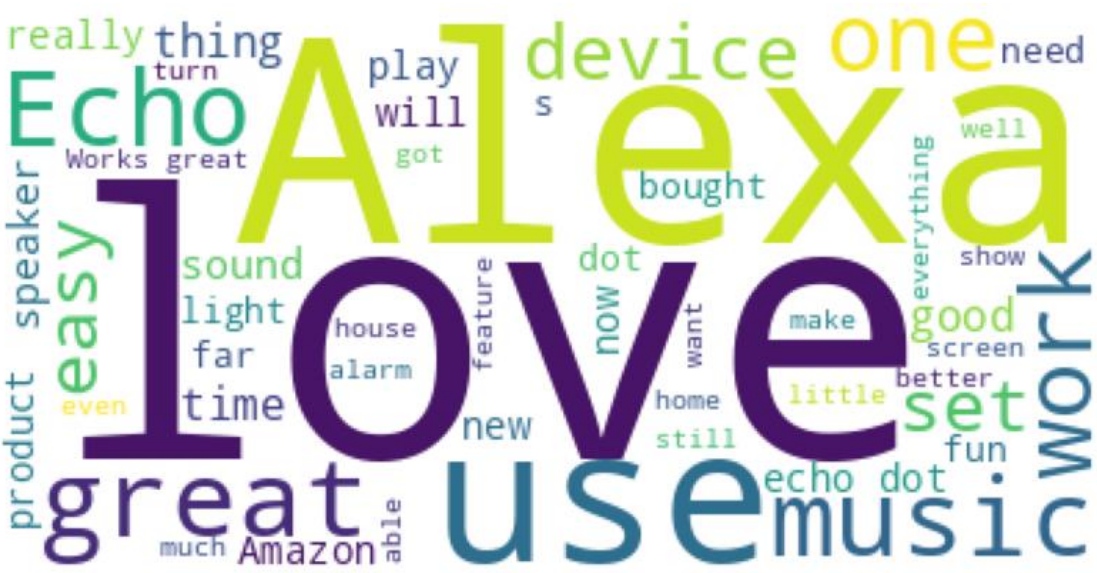


**Figure 8**: The output of customer emotions on Alexa displays the most commonly used terms from customer reviews of the Amazon electrical goods

## VII) CONCLUSION

In this study, we used the Naïve Bayes method to propose a sentiment analysis system for Amazon Alexa product reviews. We effectively categorized reviews into positive, negative, and neutral sentiments by preprocessing the review data, using feature extraction techniques like Bag of Words, and training the model. In order to better assess customer happiness and enhance product offerings, the system provides vendors and customers with insightful data. Future research should investigate more sophisticated models for even more precise sentiment classification, even if the Naïve Bayes model worked well.

## VIII) REFERENCES

1. Haque, Tanjim & Saber, Nudrat & Shah, Faisal. (2018). Sentiment analysis on large scale Amazon product reviews.10.1109/ICIRD.2018.8376299.

2. Dublin, Griffith & Joseph, Roshan. (2020). Amazon Reviews Sentiment Analysis: A Reinforcement Learning Approach. 10.13140/RG.2.2.31842.35523.

3. Tim Althoff , Cristian Danescu-Niculescu-Mizil , Dan Jurafsky Stanford University, Max Planck Institute SWS : How to Ask for a Favor: A Case Study on the Success of Altruistic Requests

4. Dey, Sanjay and Wasif, Sarhan and Tonmoy, Dhiman and Sultana, Subrina and Sarkar, Jayjeet and Dey, Monisha(February 2020) A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews.

5. Najma Sultana &Pintu Kumar & Monika Rani Patra & Sourabh Chandra and S.K. Safikul Alam (2019) : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.