**Bulletin of Business and Economics, 13(2), 142-153**
https://bbejournal.com
https://doi.org/10.61506/01.00309

# An Integrated Approach for Amazon Electronic Products Reviews by Using Sentiment Analysis

**Ameer Hamza[1], Kashif Bilal Majeed[2], Muhammad Rashad[3], Arfan Jaffar[4]**

**Abstract**
In our routine life, we interact a wide range of products, and frequently browse through digital media platforms to access their quality. Although the accessibility of online platforms, consumers often find it challenging to swiftly judge the quality of products on the basis of customer reviews. To cope this situation, the study addresses this problem by suggesting a machine learning-based solution to categorize product reviews. For this, we employ various machine learning techniques, including Random Forest, Naïve Bayes, Support Vector Machine (SVM), Stochastic Gradient Descent (SGD) Classifier, and Bidirectional Encoder Representations from Transformers (BERT). In our model, we incorporate pre-processing methods for prepare the dataset for training and utilize feature extraction techniques such as TF-IDF and word2vec which are then applied to different classifiers to analyze the reviews. Moreover, we conduct this study by using the Amazon Electronics category dataset, it reveals that BERT outperforms other classifiers with a performance score of 0.8896. Therefore, this technique not only streamlines the procedure of evaluating product quality but also enhances the accuracy of review classification, giving a real-world solution for consumers and businesses alike.
**Keywords:** Product Quality Assessment, Customer Reviews Analysis, Digital Platforms, Machine Learning Techniques, Model Evaluation

## 1. Introduction
In this modern world, we often browse digital media platforms to assess the quality of a wide range of products. On the customer point of view, the process of selecting the right products might be time-consuming based on customers' reviews. Then, the online marketplaces including social media and websites, have increased the consumer's access to purchase things easily through the product rating and reviews. For wise decision- making while product purchasing, we focus on the reviews category either good, negative, or neutral through sentiment analysis. In order to implement the best business practices, where customers can freely provide feedback by understanding people's emotions towards the product quality. From customer's prospective, it is difficult to read every review to determine the quality of the product. For this, technology can be used to analyse customer evaluations and feedback for tackling this issue.
In the light of Sentiment research, it is necessary to promote the growth by illuminating the real needs of customers for buying the products and the techniques of Natural Language Processing and Machine Learning can be used to analyse the sentiment of the product reviews (Brownfield et.al, 2020). The most valuable aspect of sentiment analysis is evaluating text data and obtaining opinions on these products. A lot of data is collected when using websites and social media, but it is challenging to analyse without the use of technologies. Thus, machine learning systems are able to quickly classify the feelings using these datasets (Wassan et.al, 2021). Understanding product quality requires sentiment analysis, which is made feasible by the increasing usage of machine learning algorithms on text input by a variety of users. Utilising Natural Language Processing, the model is developed to apply various methods to the test dataset. In 2019, Levent et al. suggested that the feelings in Amazon product reviews could be categorised using machine learning techniques. Moreover, Big Data technologies are essential for handling and transferring the vast amount of data gathered from e-commerce websites to machine learning algorithms for classification. Through Sentiment analysis a variety of Machine Learning (ML) methods, including Naïve Bayes, SVM, Decision Tree, and others are used.
The E-commerce platforms including Amazon, eBay, and Walmart are the best marketplaces for online shopping and these platforms develop confidence in customers for endorsing and reviewing the product through Sentiment analysis. However, the dataset obtained from these websites is in large volume, and is not executing without Big Data technologies; for the exploratory analysis, we can use few data that fulfill the model requirements. The dataset used in this research study was taken from text-based e-commerce websites and Natural language processing techniques were employed to aid in the analysis of the data, and machine learning algorithms were used to classify reviews as positive, negative, or neutral. The reviews of electronics products on the Amazon website are the main subject of our research study (Levent et al., 2019). The greatest places to find product reviews with datasets are Amazon and eBay. The websites of electronic devices also include thoughts, which can be examined for attitudes through the use of natural language processing and machine learning algorithms.

### 1.1. Research Objectives
We will provide the sentiment analysis solution for the dataset of reviews of Amazon Electronics products. A few important, crucial points for our research study's objectives are fulfilled by this study.
- To determine the sentiment analysis components found in customer reviews and feedback on Amazon's electronic products.
- To ascertain the sentiment's polarity, as indicated by the first objective.
- To assess the model's accuracy in identifying the sentiments.

### 1.2. Research Questions
We will address the following research topics that our proposed research model can address based on the research gaps. The following is a list of the research questions are given below:
- What feelings do Amazon Electronics Products represent?
- How are reviews and comments categorised on Amazon Electronics products?
- What are the top-performing machine learning algorithms on the Amazon Electronics Dataset?

[1] Superior University, Lahore, Pakistan, gct.175106@gmail.com
[2] Superior University, Lahore, Pakistan, kashifbilal77@gmail.com
[3] Superior University, Lahore, Pakistan, muhammad.rashad@superior.edu.pk
[4] Superior University, Lahore, Pakistan, arfan.jaffar@superior.edu.pk

- Which specific remarks have an effect on the feedback?

The current study, by leveraging advanced NLP and ML techniques, aims to bridge the gap between vast data availability and actionable insights, ultimately enhancing the decision-making process for both consumers and businesses in the e-commerce sector.

For reader's convenience, the paper organization is as follows: Chapter 2 reviews the previous studies pertinent to the research work as a foundation stone for the study. Chapter 3 outlines the methodology, detailing the methods and techniques used. Chapter 4 presents the results and discussion, highlighting the study's findings and their implications. Finally, Chapter 5 concludes the study, summarizing the key points and offering concluding remarks.

## 2. Literature Review

In this digital era, E-commerce sites, such as Amazon, eBay, Walmart, and others, are essential online marketplaces where consumers can purchase goods easily. The trust of the customer on these commerce websites is increasing because of their high quality, prompt delivery, and convenient global product purchasing options. Only consumer reviews—which also have an impact on business practices—are the basis for this trust. As we mentioned in the introduction, sentiment analysis of the customer reviews is therefore necessary. A selection of the most recent methods for sentiment analysis of customer evaluations will be covered, but many more have already been covered in the literature review section.

The proposed approach by (Alrehili et.al, 2019) discussed how would could use an ensemble machine learning approach to classify feedback reviews to get the solution for sentiment analysis. This analysis benefits both the vendor and the customer, and customer satisfaction is essential for any successful business. In this study, it was discussed the ensemble technique, which is based on the voting notion wherein an algorithm that received the majority of votes will be considered the top performer classifier. Previously, several machine learning algorithms were utilised separately. After generating the feature vector that was sent to the classifier, the suggested model was utilised to extract the features based on intensity, such as TFIDF. SVM, Naïve Bayes, Random Forest, Bagging, and Boosting are machine algorithms that are utilised to classify feedback as positive or negative. The algorithms that perform better than others are approved and verified by the ensemble approaches. After applying it Random Forest (RF) with unigram got greater accuracy values than the other classifiers.

In 2020, Kumar et al. suggested a method for obtaining feelings about Amazon's electronics products by utilising an SVM classifier and the TFIDF features vector. Finding the best answer for the sentiment components from a sizable portion of the Amazon dataset is the issue that is being tackled. The provided text data was pre-processed using a novel method before feature extraction methods like TFIDF were used. To determine sentiments such as positive, negative, or neutral replies, the acquired feature vector is fed to an SVM classifier. Because of its methodology, the suggested framework for this study fared better than other frameworks. When clients are satisfied to buy the products, sentiment analysis for the products assists to improve business and will rely on the online commerce websites. Various techniques use machine learning approaches with the help of Natural Language Processing techniques like the lexical analysis.

A comparative research of SVM and Naïve Bayes techniques was proposed by Sanjay et al. (2020) utilising the Amazon Electronics dataset, where several algorithms were deployed for sentiment analysis of electronics products. This strategy solves the research challenge of understanding product quality because it is impractical to manually go through the vast amount of feedback data for every product. According to numerous research, the most effective methods for gathering opinions on Amazon Electronics product evaluations are machine learning and artificial intelligence. In order to obtain the relevant context, this method first collects the data and pre-processes it all. SVM and Naïve Bayes classifiers received the feature vector that was obtained from the pre-processed data after features such as TFID were removed. The proposed model was evaluated using Amazon data and outperformed under the circumstances mentioned.

A number of methods are employed, most of which rely on machine learning techniques as text data produces superior outcomes when used to construct sentiment analysis systems. This study examines the in-depth conversation surrounding sentiment analysis machine learning techniques. Machine learning was employed in many sentiment analysis situations since it provided the best answer for training the provided dataset and testing the module with the unseen dataset. The model's performance was assessed repeatedly using the same dataset train. The machine learning methods that best suited the sentiment analysis task were covered in the subsection that followed.

### 2.1. Machine Learning Algorithms

Sentiment analysis for Amazon feedback reviews heavily relies on machine learning. Numerous machine learning techniques have previously been used and examined. Algorithms for machine learning automatically accept some input values, process them, and output the results in comparison with the input. We discussed machine learning algorithms that can be applied to abstract sentiment analysis.

Table 1 shows the classification algorithms categorized as supervised and unsupervised learning approaches. The supervised Learning approach is in which we have labelled datasets, and these datasets are designed to train the algorithms to predict the actual outcome. Unsupervised Learning Approach: When we have an unlabeled dataset and apply these algorithms to analyze the clusters and patterns of the data known as the unsupervised learning approach.

**Table 1: Machine Learning Algorithms**

| Supervised Learning | Unsupervised Learning |
|---|---|
| Support Vector Machine | K-Means Clustering |
| Random Forest | Principal Component Analysis |
| Decision Tree | Anomaly Detection |
| Linear Regression | Independent Component Analysis |
| Logistic Regression | |
| k-Nearest Neighbor | |
| Neural Networks | |

In order to conduct Regression and Classification, supervised algorithms are divided into two performance groups. Since several methods have been used to classify reviews as positive or negative as well as good, bad, and neutral comments, sentiment analysis is the classification problem. SVM, Random Forest, Naïve Bayes, Logistic Regression, K-Nearest Neighbour, and Perceptron are the significant classification algorithms that work best for this issue. The rationale is that these algorithms provide continuous values that cannot be utilised in the sentiment analysis problem, while producing numerical data in large quantities for the output.

Since the labelled dataset was used for training, a variety of techniques based on supervised learning algorithms were employed for the sentiment analysis after analysis. When the dataset was tagged as necessary to use the supervised learning algorithms like Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR) etc. All those algoritshm which best fit to this problem are discussed and also it was examined in the literature review section, where other people used and suggest their future work in sentiment analysis type problems. Working of the few algorithms are discussed below in detail:

#### 2.1.1. Logistic Regression

The supervised learning strategy known as the logistic function is logistic regression. It designates a specific set of classes for the observations or analysis. It predicts the likelihood of the desired variable and classes the categorical data. Classification in binary and multiclass systems Logistics Regression performs better than the analysis problem and serves the best role. Using the dataset of Amazon product reviews, the authors (Elmurngi et al., 2018) suggested a supervised machine learning method utilising logistic regression. Online buying requires businesses where clients are satisfied, and here is where the customer confidence problem arises. To begin with, the dataset is extracted, and then pre-processing techniques are used to eliminate stop words, tokenize, and identify the words that have the most influence on the dataset. Next, combine the methods for feature extraction to acquire the feature vector. Various algorithms were applied, such as SVM, Logistic Regression, and Decision Tree, but they mentioned that Logistic Regression outperforms other algorithms.

Logistic regression is statistical-based on the probability, so in this sentiment analysis, there is a large volume reviews dataset calculated in the numerics. So this model best fits the classification between good, bad, and neutral reviews class. A survey was conducted for the sentiment analysis on amazon products and how logistic regression best fit this problem (Shah et.al, 2021).

#### 2.1.2. Naïve Bayes

The supervised learning method Naïve Bayes, which is based on the Bayes theorem, is incredibly quick in comparison to other algorithms. Different tasks in the Natural Language Processing challenge with Naïve Bayes. Using conditional probability, Naïve Bayes calculates the likelihood of an event (Ray et al., 2017).

The basic machine learning classifier Naïve Bayes uses a probability strategy for each attribute. The primary use of this algorithm is in the analysis of text data. This approach is best applied in the suggested model as this study only uses text data processing in the sentiment analysis situation. Naïve Bayes, for instance, can be applied to both directional two-layered problems and many directions in single-layered problems. Equation of the Naïve Bayes probability calculation is shown in equation (1).

$$\mathrm{P}(\mathrm{A} \mid \mathrm{B}) = (\mathrm{P}(\mathrm{B} \mid \mathrm{A}) * \mathrm{P}(\mathrm{A})) / \mathrm{P}(\mathrm{B}) \text{ ----------------------- } \quad (1)$$

Where A and B are the variables used to calculate the probability of P(A | B), this algorithm is best to get the classification based on the probability of these variables. There are bad, good, and neutral variables in this research study. This model classifies the review from these types of reviews. Comprehensive research discusses how the naïve Bayes works best for this problem (Dey et.al, 2020). Naïve Bayes is used to classifying bad, good, and neutral reviews for the large volume dataset.

#### 2.1.3. Perceptron

A component of the supervised learning approach, the perceptron algorithm uses binary classification to handle most tasks, including sentiment analysis of positive or negative remarks. The weight computation determines the class category. The four steps of a perceptron are the input values, weights and bias, net sum, and activation function. Perceptrons are single-layer neural networks. Sentiment analysis and binary classification employ a variety of techniques (Brownlee et al., 2020).

(More et al., 2020) suggested a deep learning technique that classified positive and negative feedback from Amazon product reviews by applying the LSTM, Recurrent Neural Network (RNN), and GRU algorithms. The best duty for both algorithms is to obtain the sentiments insights the Amazon products reviews.

#### 2.1.4. Support Vector Machine

The supervised technique known as Support Vector Machine (SVM) seeks to locate the hyper plane in N-dimensional space, where N is the number of features that classify the data points. Hyper planes can in two or three dimensions and can be of many different types. Sigmoid, Gaussian, Laplace RBF, polynomial, gamma, and hyperbolic kernels are among the different types of kernels available for SVM. Depending on the issue, we can employ any of them; nonetheless, the linear kernel performs better than the binary classes most of the time. When dealing with intricate problems where there are numerous classes but we are unable to discern amongst them, we employ Gaussian or Hyperbolic kernels. (Gandhi et.al, 2018).

Pre-processing methods such stop word removal from the data were applied, including feature vectorization, stemming, lemmatization, and part-of-speech tagging. After that, the feature vector was subjected to the classifier SVM using the retrieved TFIDF features. The best sentiment for the Amazon product reviews is provided by this method (Nandal et al., 2020).

For both binary and multiclass classification, Support Vector Machine (SVM) is the best method. There are numerous additional variants of SVM, however this particular challenge is focused on multiclass classification. Using SVM, the hyperplane that divides the data into distinct classes is drawn. To train the dataset, this predictive model has strong features. SVM comes in the following varieties: Sigmoid kernel, Polynomial SVM, Non-Linear SVM, SGD Classifier, and Linear SVM (Nandal et al., 2020).

### 2.2. Natural Language Processing

Computer science includes Natural Language Processing (NLP) as one of its domains. It is specifically the area of artificial intelligence that makes it possible for computers to comprehend spoken words and text so that people may interpret written words. Other technologies employed under NLP, such as rule-based modelling, computational linguistics, text translation, and many more, have also been merged with NLP. Text data is typically used in sentiment analysis. By using various approaches like lexical analysis, tokenization, lemmatization, etc., NLP assists the system in understanding the text data. In order to find insights in the text data for sentiment analysis, natural language processing (NLP) is important (IBM, 2020).

Lexical analysis is the process of converting the sentences into the word tokens, or we can say the tokenization of the text data. For example, sentiment analysis useshis pre-processing step to tokenize the text data and then apply techniques to get the feature values (Dadhich et.al, 2022).

### 2.3. Summary

The literature review component of this research study covered a variety of methodologies based on the most recent findings for the sentiment analysis of Amazon Electronics goods. Natural language processing methods including tokenization and lexical analysis, as well as a variety of machine learning algorithms. Table 3 lists all of the methods that were discussed in the literature review. The study methods suggested by these approaches are well explained in this table, along with the viewpoints that are addressed for each strategy. The state-of-the-art methods employed and the methodology for classifying reviews as good, terrible, or neutral are thoroughly outlined in Table 2. This study investigated the most recent methods and addressed their drawbacks.

**Table 2: State-of-the-art Approaches**

| Research Study | Proposed Methodology | Perspective |
|---|---|---|
| (Haque et.al, 2018) | Linear Support Vector Machine, Multinomial Naïve Bayes, Stochastic Gradient Descent, Random Forest, Logistic Regression, and Decision Tree. | Six classifiers used, feature extraction techniques are TFIDF, a bag of Words, and Chi-Square, to sentiment analysis using Amazon dataset. |
| (Nguyen et.al, 2018) | Supervised Learning SVM, GB, and LR, lexicon-based approach VADER, SentiWordNet, Valence Aware Dictionary, and Pattern. | The comparative approach was applied using supervised learning algorithms and the lexical-based method which used the TFIDF features extracted to classify positive or negative feedback. |
| (Elmurngi et.al, 2018) | SVM, Logistic Regression, Decision Tree | The supervised approach TFIDF and word to vector features were extracted and used the SVM, LR, and Decision Tree. Logistic Regression plays an important role in getting the max values. |
| (Aljuhani et.al, 2019) | Naïve Bayes, Logistic Regression, Stochastic Gradient Descent and CNN. | TFIDF, Bag of Words, Glove word2vec, and word2vec with bigram features extracted and applied the Naïve Bayes, Logistic Regression, Stochastic Gradient Descent, and CNN algorithms to classify the feedback. Logistic regression gets the best results. |
| (Alrehili et.al, 2019) | Ensemble Machine Learning Approach | Amazon data is used to extract the TFIDF features, and apply the ensemble approach to get the best results such that SVM, Naïve Bayes, Random Forest, Bagging, and Boosting. SVM gets the highest values against the given dataset. |
| (Kumar et.al, 2020) | SVM Classifier | TFIDF features were extracted and feedback classified using an SVM classifier. The proposed framework used the best pre-processing approach. |
| (Sanjay et.al, 2020) | SVM and Naïve Bayes | Large volume data is used to extract the features and apply the SVM and Naïve Bayes approach where SVM gets good as compared to the Naïve Bayes classifier. |
| (More et.al, 2020) | Deep Learning RNN, LSTM, and GRU | The deep learning approach was used such that the LSTM algorithm (Recurrent Neural Network RNN) and GRU algorithm were used to classify the positive and negative feedback using Amazon products reviews. Both algorithms play the best role to get the sentiments insights the Amazon products reviews. |
| (Nandal et.al, 2020) | SVM Classifier | They applied the pre-processing techniques such that feature vectorization, part-of-speech tagging, stemming, and lemmatization, and removed the stop words from the data. Then extracted TFIDF features, the classifier SVM was applied to the feature vector. This approach gives the best sentiment for the Amazon product reviews. |

Based on the literature review, the research study looked at how sentiment analysis for Amazon products uses Support Vector Machine, Naïve Bayes, Random Forest, and Logistic Regression. There were other methods that talked about using these classifiers, but some of them didn't perform the tuned parameters. The algorithm from this list and the dataset—Amazon Electronics Products—were chosen for this research investigation. The dataset including reviews of Amazon products employs a variety of applied techniques. The Amazon dataset using SVM, Logistic Regression, and Naïve Bayes classifiers will be covered in the next section.

## 3. Methodology

This chapter discusses the quantitative approach, as we discussed various approaches applied in the literature review. There are some essential techniques addressed for the sentiment analysis. The first approach discussed the O.S.E.M.N (O: Obtained the data, S: Scrub the data, E: Explore the data, M: Model data, N Interpret data) proposed by the (Wiggins et.al, 2010), the second approach (Roy et.al, 2020) in which a bag of word approach mentioned where we tokenize the dataset, create a bag of words, extract the features, apply classifiers and finally evaluate the model using any evaluation measures. The third approach

(Symeonidis et.al, 2018) is that in which we focused on five major point necessary to implement for the sentiment analysis such as data collection, pre-processing, machine learning model for the clarification, evaluation of the model, and visual results. Based on the analysis of the previously applied techniques discussed in the literature review. Three techniques that have already been used for sentiment analysis are also described here. For the sentiment analysis, Various used the Amazon dataset. Other platform datasets from eBay, Walmart, Alibaba, and other sources were used in a variety of ways. Few of them also apply sentiment analysis to social media data. The sentiment analysis of electronics products is the foundation of our problem, and the Amazon dataset is our choice. We put out our method for doing sentiment analysis on the dataset of Amazon Electronic Products. Figure 3.1 illustrates the specifics of the suggested paradigm and explains each step in detail below. Our method is machine learning with natural language processing and feature extraction, and we concentrated on the Amazon dataset.

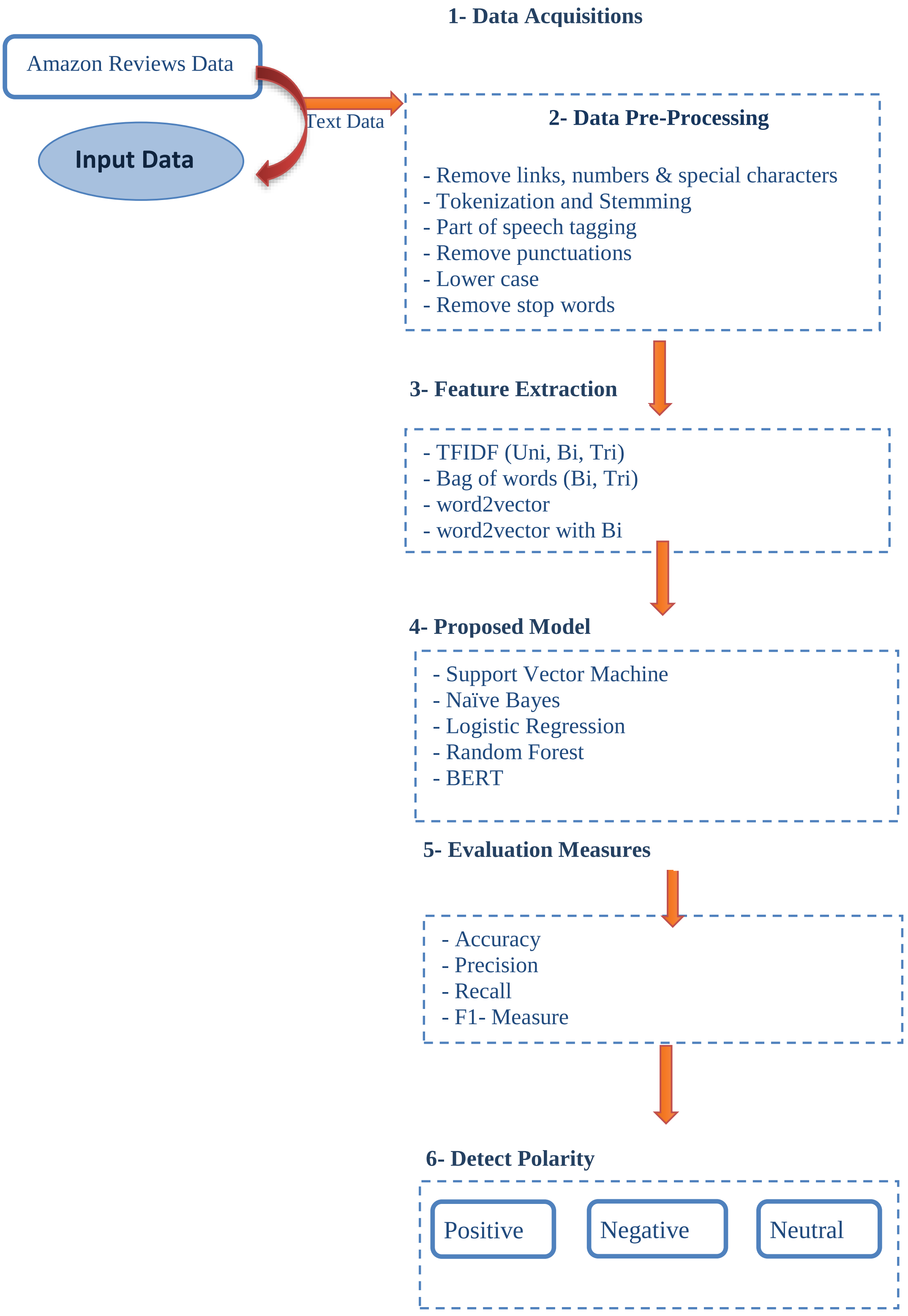


**Figure 1: The proposed architecture of the proposed model**

### 3.1. Dataset Acquisition

While several open-source datasets are available, we concentrated on the Amazon dataset. The dataset we used from Amazon is the "Amazon Review Data (2018)" (Jianmo et al., 2019). This dataset includes reviews with ratings, text, votes, and helpfulness. It is an updated version of the Amazon Review 2014 data. In addition, it comprises metadata with brand, price, category, and product image descriptions. There are links to examine the graphs and photos in the Amazon review data. Amazon reviews data comes in a variety of forms, from which we can select the ones we need, limiting our requirements to the electronics dataset. Reviews from May 1996 to October 2018 are included in the dataset reviews. Even if there is a lot of data here, it can be used for experiments.

Metadata has some specific information that is updated accordingly to product information, such as color (white or black), package type (electronics or hardcover), size (small or large), and product images. In this data, they also added more than five categories of data. For example, electronics data for the complete reviews are 20,994,354 reviews, and in metadata 786, 868 products. For the small data of electronics products reviews, data is 6.739, 590 reviews and metadata 20, 994, 353 rating data.

Data format are in one review per line in the JSON, and the given sample image is shown below:


```
{
"image": ["https://images-na.ssl-images-
amazon.com/images/I/71eG75FTJJL._SY88.jpg"],
"overall": 5.0,
"vote": "2",
"verified": True,
"reviewTime": "01 1, 2018",
"reviewerID": "AUI6WTTT0QZYS",
"asin": "5120053084",
"style": {
        "Size:": "Large",
        "Color:": "Charcoal"
        },
"reviewerName": "Abbey",
"reviewText": "I now have 4 of the 5 available colors of this
shirt... ",
"summary": "Comfy, flattering, discreet--highly recommended!",
"unixReviewTime": 1514764800
}
```

```
{
  "reviewerID": "A2SUAM1J3GNN3B",
  "asin": "0000013714",
  "reviewerName": "J. McDonald",
  "vote": 5,
  "style": {
    "Format:": "Hardcover"
  },
  "reviewText": "I bought this for my husband who plays the piano.
He is having a wonderful time playing these old hymns.  The music  is
at times hard to read because we think the book was published for
singing from more than playing from.  Great purchase though!",
  "overall": 5.0,
  "summary": "Heavenly Highway Hymns",
  "unixReviewTime": 1252800000,
  "reviewTime": "09 13, 2009"
}
```


**Figure 2: Amazon data format sample (Jianmo et.al, 2019)**

### 3.2. Data Pre-Processing

The gathered dataset will not exist in the format that we must process. Because the raw data contains undesirable text and noise, we will utilise a variety of pre-processing techniques. Various multiple techniques will be used, such that:

- Remove links, numbers & special characters
- Tokenization and Stemming
- Part of speech tagging
- Remove punctuations
- Lower case
- Remove stop words

Without preprocessing, data will compromise the accuracy value of the proposed model. So after applying these techniques, we will pass this data to the next phase, such as feature extraction.

### 3.3. Feature Extraction

All model analysis, model execution, and processing will rely on the values of the feature vectors. In order to carry out any machine learning operation, features are essential. There are two ways we will apply feature engineering in machine learning. The first involves employing methods to gather the features like:

- TFIDF (Uni, Bi, Tri)
- Bag of words (Bi, Tri)
- word2vector
- word2vector with Bi

### 3.4. Proposed Model

At this stage, Currently, we will train our model using the provided open-source dataset using the machine learning algorithm. Machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression, Random Forest, BERT, and Naïve Bayes are utilised for model training. After training the model on the provided dataset, we will assess our suggested model using the open-source dataset that is currently available. After reviewing several current methods for the sentiment analysis of text data, this study selected these algorithms because they differ from the other algorithms in a few key ways. For example, Naïve Bayes is the best classification technique, SVM selects the best hyper plane in the feature vector, and Logistic Regression is the classifier that employed probability to identify the desired aim.

#### 3.4.1. Evaluation Measures

The open-source Amazon dataset is used to assess the suggested model. Assessing the consistency and performance of the suggested model, we compute evaluation metrics including accuracy, precision, recall, and F1 Measure. This assessment method demonstrates that our suggested model performs better than the provided method. Accuracy can be defined as the ratio of accurately defined predictions to all of the predictions. The accuracy metric is displayed in Equation (2). These evaluation measures are described in terms of TP (True Positive), TN (True Negative), FN (False Negative), and FP (False Positive).

$$Accuracy = \frac{\mathrm{TN+TP}}{\mathrm{TP+TN+FP+Fn}} \quad \text{----------} \ (2)$$

Precision is the ratio of true positive and true positive, false-positive response. The equation of precision is shown in equation (3).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad \text{----------} \ (3)$$

#### 3.4.2. Detect Polarity

At the end of the proposed model classify the reviews/feedback as positive, negative, and neutral. And this is based on all previous phases and the trained model only classifies and mentions.

### 3.5. Summary

The idea behind to the choose the proposed model is it gives a clear and understandable picture of all the activities which we can perform and get the optimum solution. This proposed model elaborates each activity clearly and gives the best understanding to implement the model using the Amazon dataset.

## 4. Results and Discussion

All of the experimental findings, together with data and illustrations for the suggested model, are covered in this chapter. The model approach that was suggested in the previous chapter proposal model has been put into practice. In the process of developing the model, a few experimental data were used to validate its performance and add variables that addressed the problem statement. There are portions of the suggested methodology that we address appropriately; for example, we used the methods to obtain the necessary dataset for our model during the dataset pre-processing stage. After removing anything superfluous from the dataset, the resulting dataset was utilised for additional processing. This chapter's many sections expound on the outcomes derived from the model that is suggested below.

### 4.1. Dataset Analysis Visual Results

In order to fully understand the dataset's insights, dataset analysis is essential. We employed the models that best suit the needed dataset based on the exploratory data analysis. Figure displays the sentiment class distribution for the given dataset, including the neutral class and the distribution of positive and negative sentiment classes. These three classes are ours, and we used machine learning methods to perform categorization based on these classes. There are more positive class tweets than neutral or negative ones. The training dataset has to be balanced in order for the labelled dataset to produce accurate classification results, hence the entire dataset had to be balanced.

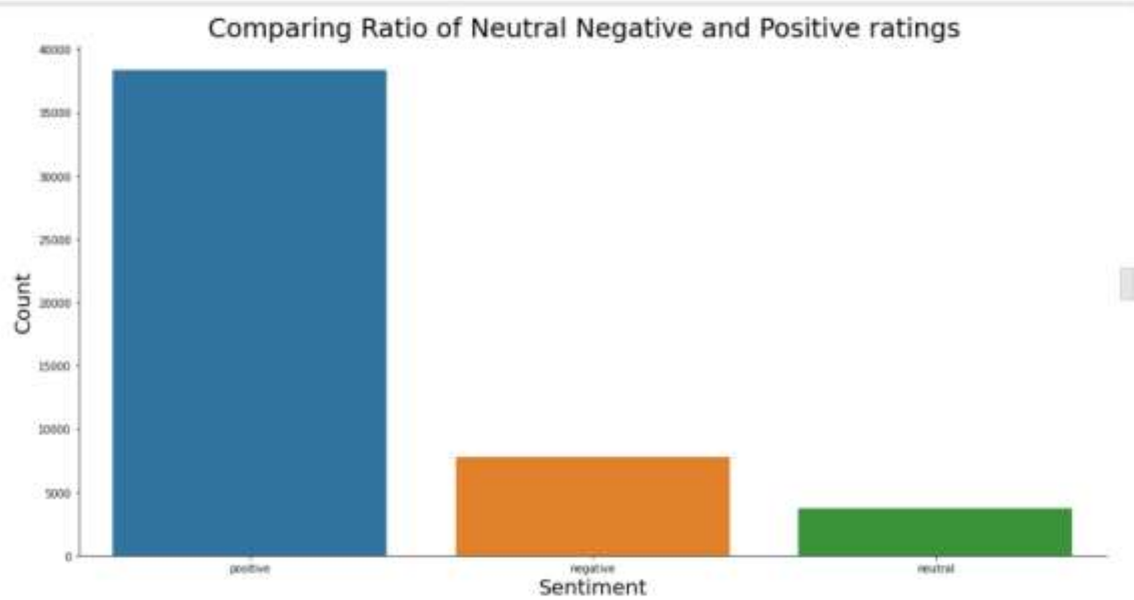


**Figure 3: Distribution of sentiment classes across the dataset**

We explored and represented the most repeated words in the dataset of neutral, positive, and negative reviews. Figure shows the most repeated words in the neutral reviews using the word cloud approach. Where we can see the neutral nature words collected and the most repeated words are bold, such as job, shape, Work, etc. Other words which were used normal range are shown respectively.

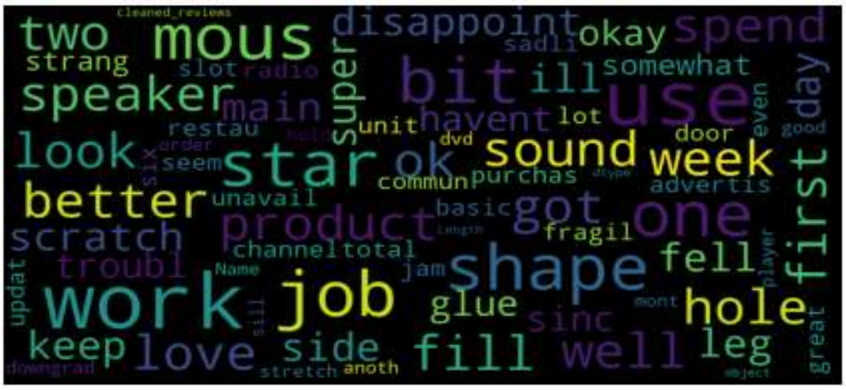

**Figure 4: Most repeated words in neutral reviews**

Words that were most repeated in positive reviews are shown in figure . Where most repeated positive nature reviews included bold words such that good, use and price, etc.

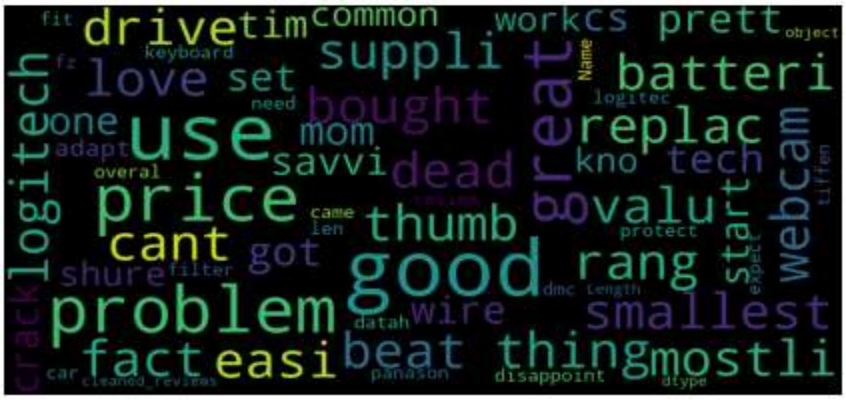

**Figure 5: Most repeated words in positive reviews**

The last most repeated words shown in the negative reviews were highlighted using word cloud in figure. In this scenario, the primarily used negative reviews for the electronics products are purchase, one and two, etc.

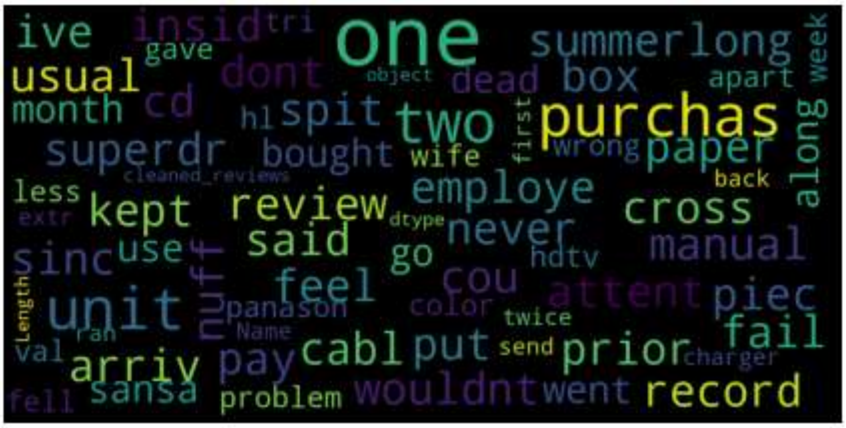

**Figure 6: Most repeated words in negative reviews**

After that, we identified the most popular words that greatly impact the overall dataset. These words can change the sentiments based on the majority voting; we identify for the bigram and remember the most valuable and major contributor words were determined shown in figure 4.5. In Good reviews bigram it shows Work great is the very popular word, in neutral reviews bigram work fine and the bad reviews bigrams stop Work is the most popular and so on.

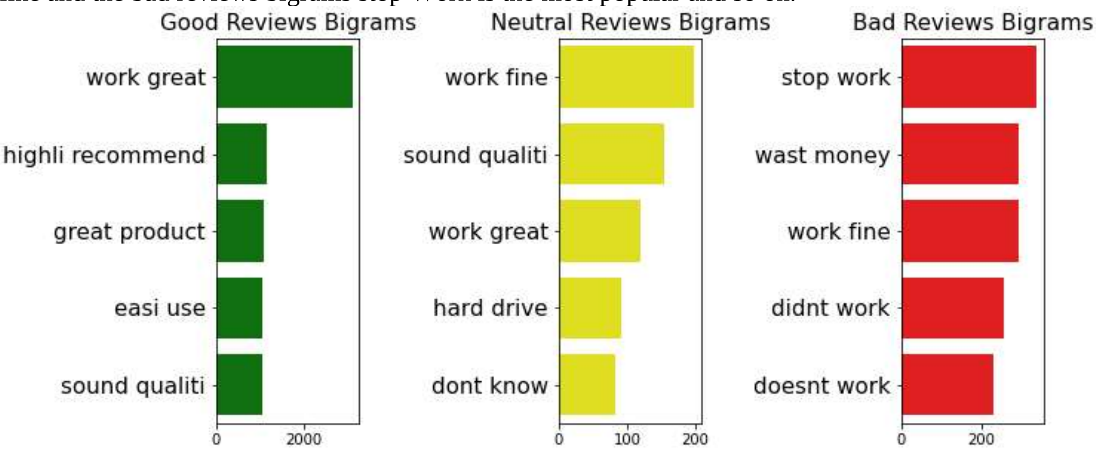

**Figure 7: Most popular words in the dataset**

Figure 4.6 shows the heat map for the correlation to show the dependency between review length and sentiments. After seeing the visuals, we conclude that we have more reviews for the neutral, and people bother themselves when they like the products and scarce chances they dislike and gave the bad reviews.

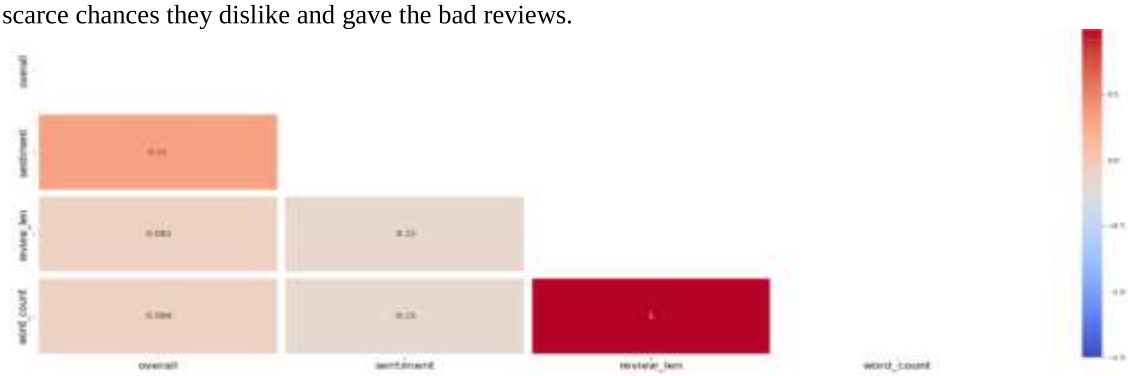

**Figure 8: Heat map from overall rating, review length, and word count**

It also shows the overall contribution of the reviews, sentiments, review length, and word count. These visual results gave us values on how we can implement the models and which model will be best for this data type.

### 4.2. Proposed Algorithm Results

Several machine learning techniques were employed in the suggested model's construction. With their optimised attributes, the Random Forest classifier yields accuracy score values of 0.823914; information gain criteria is set using entropy and the Gini index, which yields better results when using the Gini index; no processing is done at no cost; and multiple estimators or multiple trees are used, all of which add up to 100,000.

To achieve the best results, alpha values of 11 were utilised for the tokenization counter vectorizer illustrated in figure . Naïve Bayes methods were used with modified parameters so that learning rate variants were 0.0001,0.001,0.01,0.1,1,3,5,8,11,13, and 15. The Multinomial Naïve Bayes approach yielded values of 0.8254.

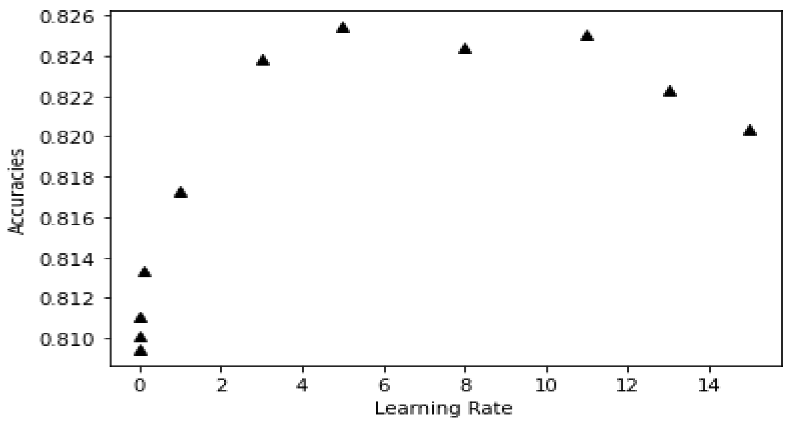


**Figure 9: Accuracy values against learning rates in Multinomial Naïve Bayes**

The results compare to the other Naïve Bayes methods such that complement Naïve Bayes results are 0.81536 and for the Bernoulli Naïve Bayes results are 0.751678. The best results in Naïve Bayes are obtained using Multinomial Naïve Bayes method.

In the Stochastic Gradient method, the learning rates are tuned with some different values such that 0.000001, 0.000003, 0.000004, 0.000006, 0.000008, 0.000009, 0.00001, 0.000012, 0.000013, 0.000014, 0.000015, 0.000016, 0.000017, 0.000018, 0.00002, 0.00003. But the results with some better values are obtained using alpha value 0.00001 using TFIDF feature extraction for the tokenization. The results obtained using Stochastic Gradient classifier are 0.84591, and the remaining results for different learning rates are shown in the figure.

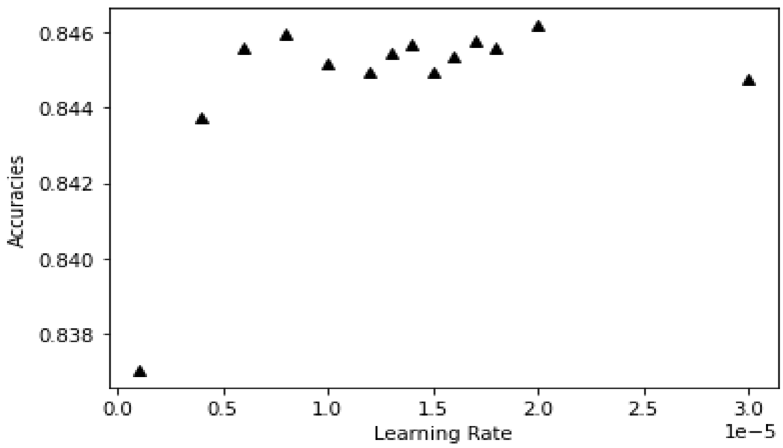


**Figure 10: Accuracy values against learning rates in Stochastic Gradients**

Naïve Bayes methods are used for the model evaluation with the tuned parameters with learning rate variations are 0.0001,0.001,0.01,0.1,1,3,5,8,11,13,15 and the best results obtained using alpha value 0.1 for TFIDF for the tokenization.

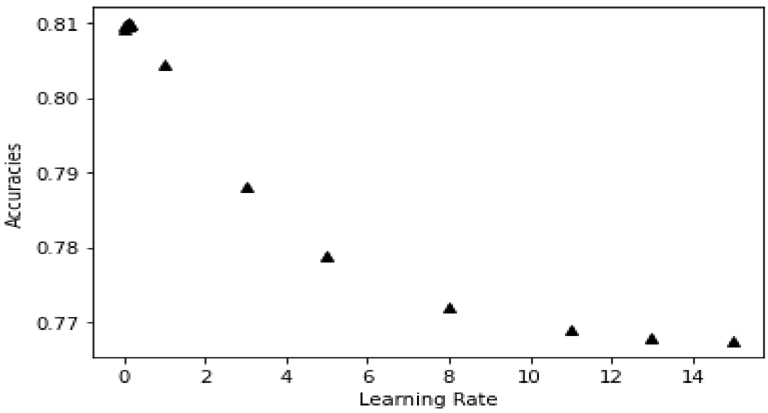


**Figure 11: Accuracy values against learning rates in Naïve Bayes**

In the end, we used the BERT algorithm, which gives the overall high values results with an 89% accuracy value. This proves that the Bert outperforms as compared to the other proposed algorithms for the sentiment analysis.

### 4.3. Classifiers Results Using Count Vectorizer

Table 4-1 displays the outcomes of the several classifiers that were employed for the model classification. The Count Vectorizer for every classifier serves as the foundation for these outcomes for the specified classifiers. Table displays the accuracy scores achieved for each classifier, along with the count vectorizer, chi, and best values. The features were first extracted using a count vectorizer, and then the various classifiers were fed these features. Finding the sentiments' classification into the three categories of neutral, positive, and negative was the aim. For instance, how does the system suggest more things to a customer who has already purchased anything? It accomplishes this by looking at the customer's past remarks on electronics products. Comments helped the retailers know the shopper behavior on the selected items they want most of them.

**Table 3: Comparative results of the classifiers with Count Vectorizer**

| No | Classifier Name | Accuracy |
|---|---|---|
| 1 | Random Forest | 0.8231 |
| 2 | Multinomial Naïve Bayes | 0.8250 |
| 3 | Complement Naïve Bayes | 0.8153 |
| 4 | Bernoulli Naïve Bayes | 0.7516 |
| 5 | SGDClassifier SVM | 0.8271 |
| 6 | SGDClassifier Logistic Regression | 0.8295 |

We extracted the TFIDF features after the count vectorizer features. Using the feature vector that we had extracted from the TFIDF, we fed it to various classifiers, including the Random Forest, Multinomial, Complement, Bernoulli, and SGD classifiers; additionally, we fed it to the SGD classifier Support Vector Machine and the SGD classifier Logistic Regression. Tables 4-2 display the statistical findings from these classifiers. While each of these classifiers gave the best results, the SGD Classifier Logistic Regression classifier did better than the others. The user comments on the electronics products sold on Amazon help this algorithm determine which of the three class feelings is most prevalent.

**Table** Error! No text of specified style in document.**: Comparative results of the classifiers with TF-IDF Vectorizer**

| No | Classifier Name | Accuracy |
|---|---|---|
| 1 | Random Forest | 0.8265 |
| 2 | Multinomial Naïve Bayes | 0.8095 |
| 3 | Complement Naïve Bayes | 0.7743 |
| 4 | Bernoulli Naïve Bayes | 0.7415 |
| 5 | SGDClassifier SVM | 0.8462 |
| 6 | SGDClassifier Logistic Regression | **0.8472** |

Results are improved using TF-IDF Vectorizer compared to the Count Vectorizer as shown in Tables 4-1 and 4-2. So SGDClassifier Logistic Regression is higher values as compared to all other classifiers.

### 4.4. BERT Classifier Results

As we used many classifiers to train, test the proposed model and also evaluate the proposed model. We experiment with the BERT classifier on the same dataset and the results obtained using BERT are very high compared to the previous classifiers, as shown in table 4-3.

**Table 5: Comparative results for BERT and other Classifiers**

| No | Classifier Name | Accuracy |
|---|---|---|
| 1 | Random Forest | 0.8265 |
| 2 | Multinomial Naïve Bayes | 0.8095 |
| 3 | Complement Naïve Bayes | 0.7743 |
| 4 | Bernoulli Naïve Bayes | 0.7415 |
| 5 | SGDClassifier SVM | 0.8462 |
| 6 | SGDClassifier Logistic Regression | 0.8472 |
| **7** | BERT Classifier | 0.8893 |

The BERT classifier's results demonstrate The BERT classifier's results demonstrate that it outperforms the count vectorizer and TF-IDF vectorizer terms classifiers. We employed three epochs in the BERT, with the outcomes for each epoch as follows: in Epoch 1, the F1-Score is 0.8756, the validation loss is 0.4041, and the training loss is 0.4641. The F1-Score was 0.8872, the validation loss was 0.4716, and the training loss was 0.3426 in Epoch 2. The F1 Score was 0.860, the validation loss was 0.5703, and the third epoch training loss was 0.2473. As we can see, values get better with the number of epochs increases. These three epochs were all that were required for mode training in this experiment, and we used them to outperform other classifiers. Therefore, the overall proposed model performance is high, and we obtained the goal based on the problem statement mentioned.

### 4.5. Comparative Analysis with State-of-the-art

The comparison analysis, which contrasts cutting-edge methods for analysing system performance, increased the suggested system's worth. This increases our incentive to develop a cutting-edge technology for sentiment analysis of electrical products. The suggested system is compared based on various factors, such as the method or methodology that was utilised to provide the answer, the kind of dataset that was used to assess the system, whether it was an open-source or privately held dataset, and the values of some assessment metrics. Tables show the proposed system's comparative analysis with state-of-the-art approaches.

**Table 6: Comparative analysis with state-of-the-art**

| Reference | Dataset | Approach | Evaluation |
|---|---|---|---|
| (Shrestha et.al, 2019) | Amazon | SVM and RNN | Accuracy 0.8125 |
| (Ganesh et.al, 2020) | Amazon | LSTM and LSTM-GRu | Accuracy 0.7235 |
| (Aljuhani et.al, 2019) | Amazon | Logistic Regression, Naïve Bayes, CNN | Accuracy 79.91 |

### 4.6. Discussion

Customer behaviour analysis is greatly influenced by product reviews on Amazon for electronics. Both good and negative product reviews have meaning in the industry, which is important to know in order to manage the finest business. It is possible to determine whether a customer has expectations based on bad customer evaluations, which can occasionally reveal the true needs of the customer. Positive product reviews offer insights into consumer behaviour and enable the addition of additional in-demand products. On the same panel, consumers rank the items that offer the finest analysis; low ratings correspond to un favourable evaluations, and high ratings correspond to favourable reviews for these products.

We performed an excellent analysis of the reviews, both favourable and negative, for the electronics products. Additionally, look up their rating to get a sense of how satisfied customers are. The suggested system uses a variety of techniques to determine how satisfied customers are with these products. A comparative examination reveals that the suggested system method works better than these cutting-edge techniques. The system's implementation of the suggested technique improves the performance of the electronics products' sentiment analysis. Because the machine learning algorithms' suggested tactics yield more positive evaluations than other products, they are better at identifying the high-selling items in Amazon's electronics category. Evaluation metrics display the performance of the suggested model using machine learning methods for various parameter values.

## 5. Conclusion

It was concluded that the solution for sentiment analysis in electronics products was essential. As part of our daily routine, we examine a variety of things and swipe through a lot of digital information, but we are unable to swiftly assess the quality of the products. Customers found it challenging to choose the intended results, thus opinions about the products were needed based on their feedback. Digital platforms such as social media, websites, and the like make it easier for consumers to access things online, purchase them, and rate and review them. We were aware that reviews may be easily categorised as excellent or poor based on rating. On the other hand, they had to examine every word in the reviews to determine the calibre of the goods.

Therefore, it was necessary to gather opinions based on what customers had to say about the products. Customers find it difficult to read lengthy evaluations and quickly determine the quality of the goods. Using machine learning techniques like Random Forest, Naïve Bayes, SVM, SGDClassifier, and BERT, we suggested a solution to this issue. The suggested methodology is able to evaluate products in real time based on the opinions expressed by users.

In the future, we will be able to conduct sentiment analysis for a variety of products using a number of different datasets. This system can be constructed using a big volume of datasets that we are unable to employ because of high processing costs while utilising various deep learning techniques for optimisation. This system can be implemented in real-time retailer shops to analyze user behavior for the other products.

### 5.1. Limitations

The resources available to do the analysis on the large data set were limited, thus this study only looked at Amazon electronics products. Sentiment analysis carried out on large data sets utilising machine learning and natural language processing algorithms ought to provide far better consumer insights and aid stakeholders in making decisions based on findings. Additionally, free text fields for reviews positive or negative allow reviewers to use slang or abbreviations, which might produce biassed findings.

Another significant flaw found in the research is the dataset polarity ratio. When the data was segmented, it was found that most evaluations were more positive than negative, which is another sign of biassed conclusions from an unbalanced sample in the learning dataset.

### 5.2. Recommendations for Future Work

The large amount of data provided by this study can be utilised to train deep learning models and create a more comprehensive model once a variety of factors have been examined. This model has the potential to produce more accurate results and resolve more challenging sentiment analysis issues. In addition, a better pre-processing of the model could optimise its parameters. Best-fitting parameters can be used to train deep learning models for later use, and this application can be used for profit.

## References

Aljuhani et.al. (2019). A comparison of sentiment analysis methods on Amazon reviews of Mobile Phones. *10*, 608--617. Retrieved 11 25, 2021

Aljuhani et.al. (2019). A comparison of sentiment analysis methods on Amazon reviews of Mobile Phones. *Int. J. Adv. Comput. Sci. Appl, 10*, 608--617. Retrieved 4 22, 2022

Alrehili et.al. (2019). Sentiment analysis of customer reviews using ensemble method. *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1--6). IEEE. Retrieved 11 24, 2021

Brownfield et.al. (2020). Sentiment Analysis of Amazon Product Reviews. *Proceedings of the Computational Methods in Systems and Software* (pp. 739--750). Springer. Retrieved 11 25, 2021

Brownlee et.al. (2020, 12 11). *Perceptron Algorithm for Classification in Python*. Retrieved 11 26, 2021, from machine learning mastery: https://machinelearningmastery.com/perceptron-algorithm-for-classification-in-python/

Dadhich et.al. (2022). Sentiment analysis of amazon product reviews using hybrid rule-based approach. In A. Dadhich, *Smart Systems: Innovations in Computing* (pp. 173--193). Springer. Retrieved 6 23, 2022

Dey et.al. (2020). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. In S. Dey, *2020 International Conference on Contemporary Computing and Applications (IC3A)* (pp. 217--220). IEEE. Retrieved 6 23, 2022

Elmurngi et.al. (2018). Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques. *J. Comput. Sci.*, 714--726. Retrieved 11 26, 2021

Gandhi et.al. (2018, 6 7). *Support Vector Machine — Introduction to Machine Learning Algorithms*. Retrieved 11 26, 2021, from towards data science: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Ganesh et.al. (2020). Sentiment Analysis on Amazon Product Reviews with Stacked Neural Networks.

Haque et.al. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development* (pp. 1--6). Bangkok, Thai: IEEE. Retrieved 11 26, 2021

IBM. (2020, 7 2). *What is Natural Language Processing?* Retrieved 11 26, 2021, from IBM: https://www.ibm.com/cloud/learn/natural-language-processing

Jianmo et.al. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In J. a. Ni, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 188--197). Retrieved 11 26, 2021

Kumar et.al. (2020). A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus. *2020 National Conference on Communications (NCC)* (pp. 1--6). IEEE. Retrieved 11 23, 2021

Levent et.al. (2019). Sentiment analysis for amazon. com reviews. *Big Data in Media Technology (DM2583) KTH Royal Institute of Technology, Stockholm*. Retrieved 23 11, 2021

More et.al. (2020). Sentiment Analysis on Amazon Product Reviews with Stacked Neural Networks. *no. October*. Retrieved 11 26, 2021

Nandal et.al. (2020). Machine learning based aspect level sentiment analysis for Amazon products. *Spatial Information Research, 28*, 601--607. Retrieved 11 26, 2021

Nandal et.al. (2020). Machine learning based aspect level sentiment analysis for Amazon products. *Spatial Information Research*, 601--607. Retrieved 6 23, 2022

Nguyen et.al. (2018). Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches. *SMU Data Science Review, 1*, 7. Retrieved 11 26, 2021, from https://scholar.smu.edu/datasciencereview/vol1/iss4/7

Ray et.al. (2017, 9 11). *6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R*. Retrieved 11 26, 2021, from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

Roy et.al. (2020, 7 11). *A Guide to Text Classification and Sentiment Analysis*. Retrieved 11 26, 2021, from Towards Datascience: https://towardsdatascience.com/a-guide-to-text-classification-and-sentiment-analysis-2ab021796317

Sanjay et.al. (2020). A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews. *2020 International Conference on Contemporary Computing and Applications (IC3A)* (pp. 217--220). IEEE. Retrieved 11 24, 2021

Shah et.al. (2021). Sentiments Detection for Amazon Product Review. In B. K. Shah, *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1--6). IEEE. Retrieved 6 23, 2022

Shrestha et.al. (2019). Deep learning sentiment analysis of amazon. com reviews and ratings. *arXiv preprint arXiv:1904.04096, 8*. doi:10.5121/ijscai.2019.8101

Symeonidis et.al. (2018, 3). *5 Things You Need to Know about Sentiment Analysis and Classification*. Retrieved 11 26, 2021, from KDnuggets: https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html

Wassan et.al. (2021). Amazon Product Sentiment Analysis using Machine Learning Techniques. *Revista Argentina de Cl{\'\i}nica Psicol{\'o}gica, 30*, 695. Retrieved 11 24, 2021

Wiggins et.al. (2010, 9 25). *A Taxonomy of Data Science*.