Library Progress International *Print version ISSN 0970 1052*
Vol.44 No. 3, Jul-Dec 2024: P.14369-14382 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Machine Learning Approach to Analyze Sentiment of Customers using NLP Text Summarization

[1]Subhash Chandra Gupta*, [2]Noopur Goel

[1]Research Scholar, Department of Computer Applications, V.B.S. Purvanchal University, Jaunpur, India; csubhashgupta@gmail.com
[2]Department of Computer Applications, V.B.S. Purvanchal University, Jaunpur, India; noopurt11@gmail.com

How to cite this article: Subhash Chandra Gupta,Noopur Goel (2024) Machine Learning Approach to Analyze Sentiment of Customers using NLP Text Summarization. *Library Progress International,* 44(3), 14346-14368.

ABSTRACT
Dealers selling items online frequently request that their clients review the items that they have bought and the related administrations. As web based business is turning out to be increasingly well known, the quantity of client surveys that an item gets develops quickly. For a famous item, the quantity of surveys can be numerous. This makes it challenging for an expected client to peruse them to settle on an intellectual choice on whether to buy the products/services. It likewise makes it challenging for the business owners of the product to follow along and to oversee client reviews. For the business owner, there are extra hardships in light of the fact that numerous producers might sell a similar item and the maker regularly delivers numerous sorts of products. In this investigation, we mean to mine and to summarize all the client overviews of products. In contrast to the typical text synopsis, this rundown task is unique because we only focus on the aspects of the product that customers have commented on and whether they are optimistic or pessimistic. We don't sum up the surveys by choosing a subset or revise a portion of the first sentences from the surveys to catch the central matters as in the exemplary message synopsis. Our errand is acted in three stages: (1) mining item includes that have been remarked on by clients; (2) recognizing assessment sentences in each survey and concluding whether every assessment sentence is positive or negative; (3) summing up the outcomes. This paper proposes a few novel strategies to play out these errands. Our exploratory results using studies of different things sold internet based show the feasibility of the techniques.

KEYWORDS
NLP, Text summarization, Machine learning, Sentiment analysis, Product review, Mobile phones.

1. Introduction

Lorem With the quick development of internet business, an ever increasing number of items are sold Online, and that's just the beginning and more individuals are likewise purchasing items on the web [1]. To further develop buyer devotion and shopping experience, it has transformed into a run of the mill practice for online merchants to enable their clients to review or to give perspectives on the things that they have purchased. With an ever increasing number of normal clients becoming alright with the Internet, a rising number of individuals are reviewing the products [2]. Thus, the quantity of surveys that an item gets develops quickly. A few famous items can get many surveys at some huge trader locales. Moreover, many surveys are long and have a couple of sentences containing suppositions on the item [3]. This makes it difficult for a possible client to peruse them to go with a review on whether to buy the item. In the event that he/she just peruses a couple of surveys, he/she might get a one-sided view. The huge number of surveys additionally makes it hard for item makers to monitor client assessments of their items. For an item maker, there are extra challenges on the grounds that numerous vendor locales might sell its items, and the maker might deliver numerous sorts of items [4].
This investigation focuses on the issue of creating highlight-based outlines of customer reviews of online-sold goods (cell phones) [5]. Here, highlights comprehensively mean item elements (or traits) and capabilities. Given a bunch of client surveys of a specific item, the undertaking includes three subtasks: (1) distinguishing elements of the item that clients have offered their viewpoints on (called item includes); (2) for each component, recognizing survey sentences that offer positive or negative perspectives; and (3) creating an outline utilizing the

found data.
Assuming that a feature based dataset in which the features are given of different mobile phones, mobile phones. The summary is given as:

Mobile_ phones:
Feature: phone rating
    Positive: 293399
        (Personal review words)
    Negative: 204620
        (Personal review words)
Feature: Reviews
    Positive: 255152
        (Personal review words)
    Negative: 186754
        (Personal review words)
        ……

Fig 1: A model summary

In Figure 1, mobile phone rating and reviews of clients are the item includes. There are 293399 client evaluations that offer positive viewpoints about the mobile phone quality, and just 204620 that offer negative viewpoints. The (personal review words) interface focuses to the particular words as well as the entire words that give positive or negative remarks about the element [6].
With such a component based result, a potential client can undoubtedly perceive how the current clients feel about the mobile phone. In the event that he/she is exceptionally keen on a specific element, he/she can penetrate somewhere near following the (individual survey words) connection to see the reason why existing clients like it or potentially what they whine about. For a maker, it is feasible to consolidate results from different trader destinations to deliver a solitary report for every one of its items [7].
Our errand is not the same as customary text summarization in various ways. As a matter of some importance, an outline for our situation is organized as opposed to another (yet more limited) free text report as created by most text result frameworks. Second, we are just keen on elements of the item that clients have feelings on and furthermore whether the suppositions are good or pessimistic. We don't sum up the surveys by choosing or revamping a subset of the first sentences from the surveys to catch their primary concerns as in customary message synopsis [8]. As demonstrated over, our errand is acted in three primary advances:
(1) This examination will zero in on utilizing Natural Language methods to track down wide patterns in the composed considerations of the clients [9]. The objective is to foresee whether clients suggest the item they bought involving the data in their survey text.
(2) One of the moves in this study is to remove redundant data from the review text variable utilizing text mining strategies [10]. The other test is that we want to change over text documents into numeric element vectors to run AI calculations. At long last, we choose the assessment direction of each sentence.
(3) By applying several machine learning algorithms (Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest and Ada Boosting) we will built a sentiment analysis (SA) model which step totals the aftereffects of past advances and presents them in the arrangement of Figure 1.
In section 3 a feature based summarization has been implemented for performing above said task.

2. Literature Review

Presently, today's scientists have extraordinary interest in sentiment based review mining and text summarization for business efficiency. Accurate review analysis from large datasets is difficult because of sarcasm, noisy data, emoticons, and unreliable features.

2.1 Machine learning based sentiment analysis

For estimating E-Com administration quality in a concentrate by [11] from online client survey utilizing opinion investigation, used sentiment analysis to survey client reviews of the e-com organization Tokopedia, as shared on the site. Utilizing the Naive Bayes, they characterized the reviews into positive and pessimistic opinions for five components of electronic help quality, which incorporates website structural design, reliability, sensitivity and expectation. According to their findings, Tokopedia received high marks for trust and web design, but low marks for personalization and reliability.
In one more concentrate by [12], Opinion Investigation in Web based business utilizing Naive Bayes Technique, used feeling investigation to analyze the public feelings towards online business organizations like Bukalapak, Lazada and Tokopedia. The writer assembled the reviews from every one of the organization's Facebook pages

and used the Naive Bayes technique to arrange negative and positive remarks from each organization's clients. The aftereffects of the review uncovered that Lazada and Tokopedia had the best survey, though Bukalapak had more regrettable feelings. The Naive Bayes Classifier was exceptionally valuable in this review to arrange negative and positive remarks, which gave an effectively justifiable however precise survey of each organization in view of its particular users' experiences.

In a concentrate by [13], Opinion Investigation for Ladies' Web based business surveys utilizing AI Calculations, utilized the Weka programming of AI calculations to lead its feeling investigation of ladies' design items on Amazon, the internet business stage. The creators utilized Bayes Hypothesis and Support Vector Machines, four distinct classifications of classifiers alongside information pre-handling capabilities, emphasize extraction and property selection. The aftereffects of the review showed that SVM classifiers gave the most dependable outcomes out of the four classifiers, and the creators additionally suggested testing different calculations like CNN and KNN and sentiment examination applications.

Studies by [14], have shown the way that client opinion examination can help organizations in keeping away from negative standing and give speedy and exact solutions to business-related questions. What's more, client buying choices have been demonstrated to be affected by the feeling factors that show up in other client surveys.

As organizations get sufficiently close to a lot of information that were not open to them before, knowing how to examine and use this information has become more significant than any other time. Sentiment examination is one of the manners in which that specialists and organizations can dissect information that is introduced in a literary configuration, as opposed to an essential mathematical organization [15].

Support Vector Machines work by finding the ideal hyperplane inside a named preparing dataset, where information is plainly characterized [16]. Despite the fact that it is generally difficult to do, investigating information through feeling examination can give an abundance of data and frequently give more knowledge than fundamental mathematical information.

A few examinations have been finished in this field which investigates the various strategies for doing sentiment investigation. In a review [17], they review a few distinct techniques for sentiment examination, for example, dictionary based and observed AI. Checked AI approaches incorporate the Naive Bayes technique, which is generally utilized in the retail business, though the vocabulary based strategy uses predefined word expressions and assessment figures of speech where each expression and saying is evaluated as one or the other positive or negative opinion. Their article reasoned that sentiment examination can be an exceptionally helpful device for organizations to upgrade their client experience, but there are as yet many difficulties in executing it, for example, the utilization of mocking surveys which can be hard to characterize as good or pessimistic consequently.

In a review performed by [18], SA on Web based business Item Utilizing AI and Mix of TF-IDF and Backward Elimination, the creators applied sentiment examination to shopper item surveys in e-com stages in Indonesia with a more streamlined strategy. The strategy tried in this review was the Terms Frequency – Inverse Document Frequency (TF-IDF) technique for highlight extraction, in backward elimination for the element determination stage, and a correlation between five sorts of classifiers (NB, K-NN, DT, RF, and SVM). From the consequences of this review, it was observed that the best degree of precision was accomplished by utilizing the TF-IDF technique, in backward elimination, and afterward involving SVM as a classifier. It was additionally found that utilizing the backward elimination strategy at the component determination stage helped increment the precision of the five kinds of classifiers in leading sentiment examination.

[19], presented an ML based recommender framework for SA. Application programming connection point is utilized to gather information from the twitter information source. The gathered tweets are preprocessed and delegated positive, negative, or unbiased. At last, the viability of the three essential regulated ML calculations is assessed. The structure for SA with recommender frameworks was created utilizing supervised learning methods in view of the chose ML models.

[20], proposed an original framework for item suggestion in web based shopping, which uses fuzzy logic to progressively foresee the most fitting items for clients taking into consideration their progressing tendencies. The paper acquainted a creative strategy with process the feeling score of an item considering the particular objective class related with end clients. The proposed suggestion framework, which uses fuzzy principles and ontology based methods, integrates ontology arrangement to upgrade the exactness of decisions and make expectations that are logically significant in light of the pursuit setting. The results from the examinations demonstrated that the recently proposed suggestion framework beats the current item suggestion frameworks as far as precisely

foreseeing significant items for explicit buyers and conveying these proposals in a more proficient way. The significant constraints of the proposed work are mind boggling to create and keep up with and furthermore delicate to changes in client conduct or inclinations.

From these past examinations, there are a few investigations that utilization an AI approach with different classifiers, for example, NB, DT, SVM, and more. Despite the fact that there have been a few investigations looking at the viability of these different AI characterization strategies, frequently these examinations apply feeling investigation to a wide range of items accessible on a web based business stage or to customer evaluations of the web based business stage all in all. There have been no investigations that have zeroed in on looking at the viability of AI grouping techniques, for example, NB and SVM, on purchaser surveys of a specific sort of item. As a matter of fact, a more top to bottom investigation like this can yield more unambiguous bits of knowledge not just for online business organizations all in all, yet additionally for dealers of these items.

3. Background

In recent times, SA has received significant attention from researchers to accomplish different applications; therefore a framework to perform scalable and efficient SA is necessary [21]. We studied various methods to perform SA with the focus on effective feature representation and machine learning techniques for classification. The large number of online reviews containing spam, fake, emoticons, negation, sarcasm, etc. makes feature extraction challenging. Machine learning techniques such as supervised and unsupervised methods can be used to classify online reviews into positive or negative sentiments [22]. We believe that the accuracy of machine learning techniques mainly depends on the feature set.

By considering the above challenges of state-of-the-art methods, we propose the novel approach of SA with special reference to customer review summarization. The proposed method consists of pre-processing (feature engineering and data cleaning), text mining (tokenization, noise removal, lexicon normalization), word cloud (review detection, word collection, cloud creation) and classification phases with machine learning including train-test split, vectorization, TF-IDF and several machine learning classifiers. The pre-processing uses the common methodology to denoise the input reviews. Figure 2 depict the process of sentiment analysis through machine learning.

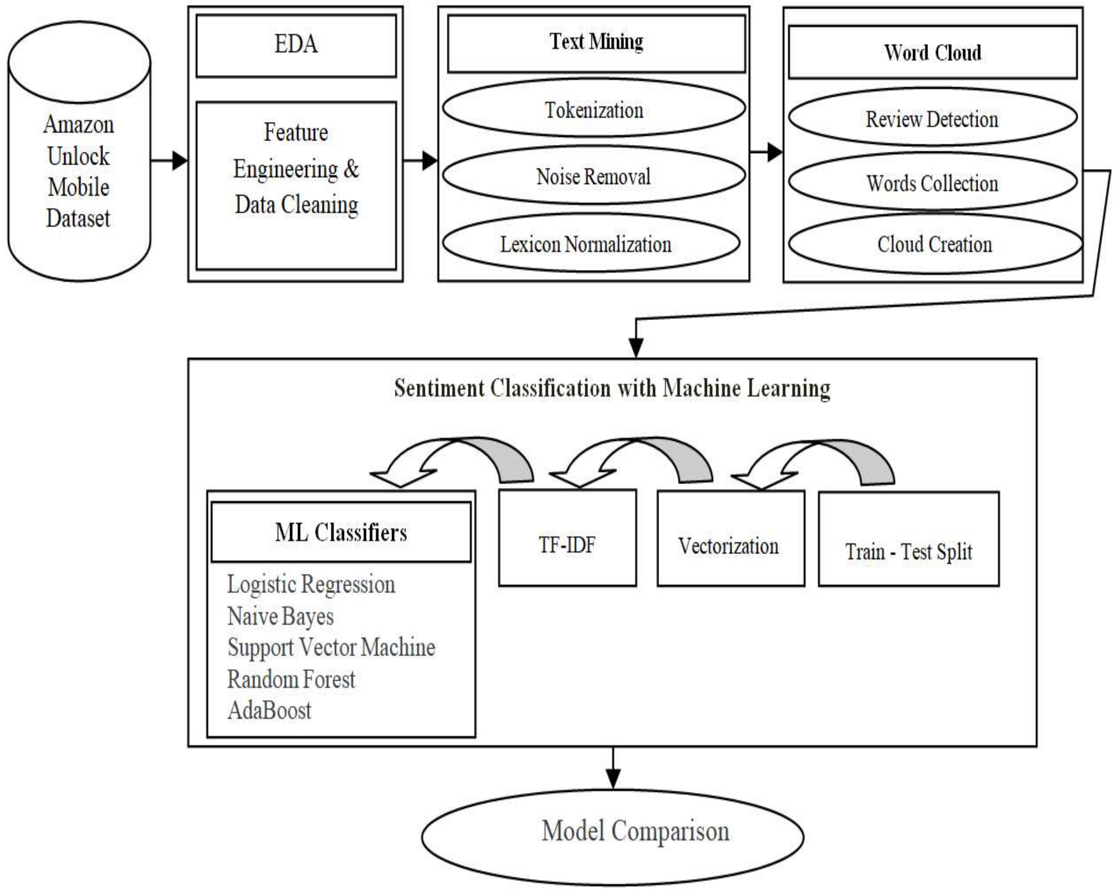


Fig. 2 Proposed frameworks for sentiment analysis

3.1 Feature engineering and Data cleaning

When searching various documents, feature engineering reveals the relevant data elements. It likewise centers on separating organized data from free text and putting away these substances, traits, and relationship data in a data set.

The data cleaning performs the cleaning of the raw reviews by removing and correcting the complex and unnecessary text [23]. The data preprocessing that starts with tokenization and ends with removing numbers and meaningless words. The tokenization function split the input review into different tokens. Then on each token, we apply stemming to reduce the tokens into their singular form. Then the stop-words like 'a', 'an', ‘and’, ‘are’, ‘as’, ‘at’, ‘be’, ‘but’, ‘by’, ‘for’, ‘if’, 'I', ‘am’ etc. are removed to reduce the number of tokens. The special characters (@, #, $, %, & etc.), dates, meaningless words (y+, Z−, etc.), and any URLs are also discovered and removed. Additionally, the algorithm also checks for the words with less than three characters and numbers and removes them.

3.2 Text Mining

The process of text mining comprises several activities that enable us to deduce information from unstructured text data [24]. Before we can apply different text mining techniques, we must start with text preprocessing, which is the practice of cleaning and transforming text data into a usable format. This practice is a core aspect of natural language processing (NLP) and it usually involves the use of techniques such as language identification, tokenization, part-of-speech tagging, chunking, and syntax parsing to format data appropriately for analysis. After the text preprocessing process is finished, we use text mining algorithms to find insights in the data. These include:

Tokenization is the process of breaking text object into a smaller unit. Tokens can be numbers, characters, words, or images. The most well-known tokenization process is whitespace tokenization. In this cycle whole text is parted into words by dividing them from whitespaces. For instance the entire sentence is parted into whitespace for example "I"," buy","a","new","phone" and so forth.

In NLP, noise can come from a variety of things, like human error, how data is collected, how it is processed, or outside factors. For example, human errors can include spelling mistakes, grammar errors, typos, slang, abbreviations, or ambiguous expressions.

The process of translating or transforming a non-standard text into a standard register is called lexical normalization.

For e.g.:

New mob launchingg tommoroe

New mobile launching tomorrow

Reviews typically make up data sets because they naturally contain a lot of these phenomena. Only word-level replacements are annotated for lexical normalization. However, the task does not include word insertion/deletion or reordering.

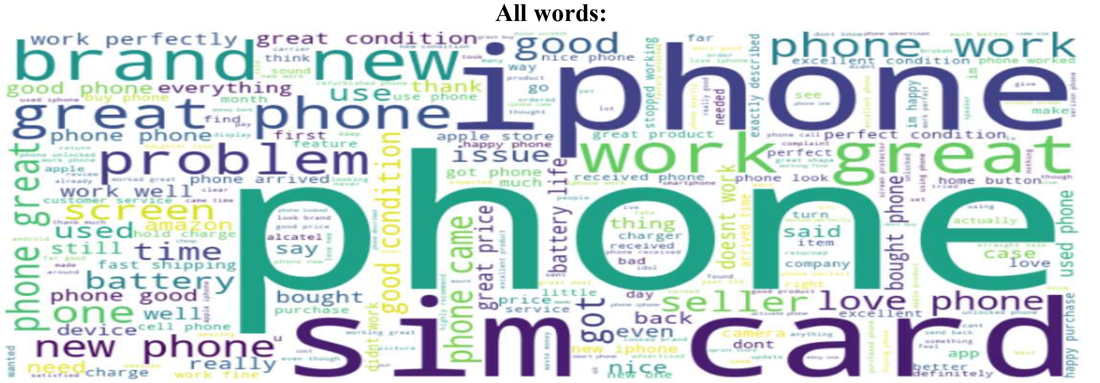
3.3 Word Cloud

Word cloud is a visual portrayal of word recurrence. The generated image shows the word in greater detail the more frequently it appears in the analyzed text. It is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance [25].

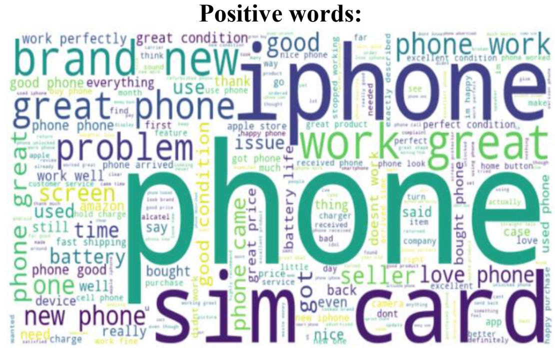
It is a cluster of keywords of varying sizes that represents the metadata in the. Word clouds can tell you which words, and thus, what elements in your data are mentioned the most and which aren't, thus giving you a direction in which to analyze your insights further. Word clouds are represented in the form of words that are formatted differently based on their relevance. Thus the more important the word, the bigger, bolder, and more centrally placed it is - thus taking the shape of a cloud. The review detection, words collection and cloud creation process are showing in the following table 1 as all words collection, positive words and negative words from our dataset review about mobile phones.

Table1. Collection of words through cloud

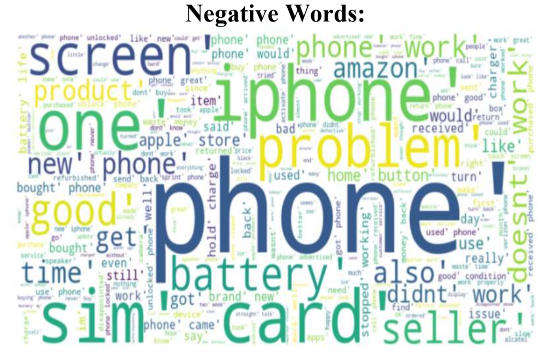
All words:



Positive words:



Negative Words:



3.4 Sentiment analysis with machine learning

SA is a machine learning tool that looks for positive and negative polarity in texts. Using textual examples of emotions as training material, machine learning tools automatically learn to recognize emotion without human intervention.

To put it simply, machine learning (ML) allows computers to learn new tasks without being expressly programmed to perform them. It is possible to train SA models to read beyond just definitions, comprehending things like sarcasm, misused words, and context.

Train-test split is a method of calculating machine learning representation. It can be used for classification or regression problems and can be used for any supervised learning algorithm [26]. The procedure involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The model is trained without using the second subset; instead, the dataset's input element is given to the model, and predictions are made and compared to the expected standards. This second dataset is referred to as the test dataset. In our case 80% dataset is consider for training and 20% for testing.

Vectorization, when we need to convert text data into the numerical data then vectorization is used in NLP [27]. It involves converting unstructured textual data into a structured format that makes it easier to better analyze and manipulate data. These techniques convert textual data into vectors of numbers, enabling machine learning algorithms to perform tasks such as classification, semantic analysis, and prediction. Vectorization has a number of benefits, including facilitating semantic analysis, enhancing textual data comprehension, and increasing the effectiveness of text-based machine learning models. It can be used for sentiment analysis, machine translation, and chatbot creation, among other things. Vectorization offers a few benefits, including working on the effectiveness of text-based AI models, empowering semantic investigation, and improving the comprehension of text based information. It can be used for a wide range of things, including chatbot development, sentiment analysis, and machine translation.

Term Frequency - Inverse Document Frequency (TF-IDF) estimates how significant a term is inside a record comparative with an assortment of reports (i.e., comparative with a corpus) [28]. Words inside a text report are changed into significance numbers by a text vectorization process. Text vectorization scoring schemes abound, with TF-IDF being one of the most widely used.

As its name implies, TF-IDF vectorizes/scores a word by multiplying the word's Term Frequency (TF) with the Inverse Document Frequency (IDF).

TF of a term or word is the times the term shows up in a record contrasted with the complete number of words in the report.

$$TF = \frac{\begin{array}{c}(No.of\ time\ the\ term\ appear\ in\\ the\ dockument)\end{array}}{(Total\ no.of\ term\ in\ the\ dockument)}$$

IDF of a term reflects the proportion of documents in the corpus that contain the term. Words unique to a small percentage of documents (e.g., technical jargon terms) receive higher importance values than words common across all documents (e.g., a, the, and).

$$IDF = \log\left(\frac{No.of\ documents\ in\ the\ corpus}{\begin{array}{c}No\ of\ documents\ in\ the\ corpus\\ having\ term\end{array}}\right)$$

A term's TF-IDF is determined by multiplying the TF and IDF scores.

$$TF\_IDF = TF \times IDF$$

ML Classifiers

As shown in Fig. 2, ML classifiers have been used to estimate the SA from the user's review. This mechanism of investigation elaborates the sentiments of end-users. As mentioned earlier, five machine learning techniques: LR, NB, SVM, RF, and Adaboost are used to perform sentiment classification in the proposed work.

LR: Based on a particular data set of independent variables, LR estimates the likelihood of an event, such as a

positive or negative vote [29].
This logit model is often used for classification and predictive analytics. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In LR, a logit change is applied on the chances that is, the likelihood of progress partitioned by the likelihood of disappointment.

NB: Naïve Bayes is part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category [30]. Unlike discriminative classifiers, like LR, it does not learn which features are most important to differentiate between classes.
Naïve Bayes is also known as a probabilistic classifier since it is based on Bayes' Theorem. It would be difficult to explain this algorithm without explaining the basics of Bayesian statistics. This hypothesis, otherwise called Bayes' Standard, permits us to "reverse" restrictive probabilities.
SVM: SVM is a supervised machine learning algorithm that uses an N-dimensional space to find the best line or hyperplane between each class to classify the data [31]. SVMs are frequently utilized in classification issues. They use the best hyperplane that maximizes the distance between the closest data points of opposite classes to differentiate between two classes. The number of features in the input data determines if the hyperplane is a line in a 2-D space or a plane in a n-dimensional space. The algorithm can determine the best decision boundary between classes by maximizing the margin between points because classes can be distinguished by multiple hyperplanes. As a result, it is able to accurately predict classifications and generalize to new data. The lines that are adjacent to the optimal hyperplane are known as support vectors as these vectors run through the data points that determine the maximal margin.

RF: RF is an ordinarily utilized AI calculation, reserved by Leo Breiman and Adele Cutler that joins the result of different decision trees to arrive at a solitary outcome [32].
The random forest algorithm is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. This is a vital contrast between choice trees and RF. While decision trees consider all the possible feature splits, random forests only select a subset of those features.

Adaboost: A meta-algorithm for machine learning is Adaptive Boosting [33]. AdaBoost is adaptive in the sense that instances misclassified by previous classifiers are favored by subsequent weak learners. Outliers and noisy data affect AdaBoost. It may be less susceptible to the overfitting issue in some situations than other learning algorithms. The individual learners may be weak.
From the study of the existing literature, we noticed that all these classifiers have several applications in text classification and have shown better performances due to their accuracy and simplicity, which encourages their use in the proposed work. For sentiment classification, the test feature vector along with the training feature vector and associated labels are given as input to each of the above-mentioned classifiers.

4. Experimental Setup

The dataset used for this research is the Amazon_Unlocked_Mobile reviews dataset [34]. The reviews in the dataset are consists of the 6 attributes such as: Product Name, Brand Name, Price, Rating, Reviews and Review Votes. Additional 7th independent attribute namely Positive rated has been created by Review votes by applying if rating is greater than 3 then it consider boolean value 1 otherwise 0. The total number of instances is 49999 for main source of data used is the product reviews from Amazon. The reviews for a few popular phones have been obtained by Amazon have been mentioned in the following table 2.

Table 2 Attribute information about phone reviews

Attribute	Description
Product Name	The name of the Product.
Brand Name	Name of the parent company.
Price	Product's price.
Rating	Rating of the product ranging between 1-5
Reviews	Description of the user experience
Review Votes	Number of people voted the review (Min: 0, Max: 645)
Positive rated	People's positive rating (If Rating>3 then 1 otherwise 0)

5. Result

For experimental analysis of the proposed model, we used the Python 3.7 as a tool under Windows 11 OS with an I5 processor and 8 GB RAM. We have used Amazon_Unlocked_Mobile reviews dataset to investigate the performance of the proposed model with state-of-the-art methods. We have taken almost 80% of 49999 instances as a training reviews and approx 20% as testing reviews of dataset.
This study aimed to predict customers' sentiments based on their online reviews. The following models were used to predict sentiments: LR, NB, SVM, RF and AdaBoost. F1_Score, Recall_ scores, and Average_Precision_Score were calculated. Correlation is a statistical measure [35]. There are three different conditions arise in the following figure 3:

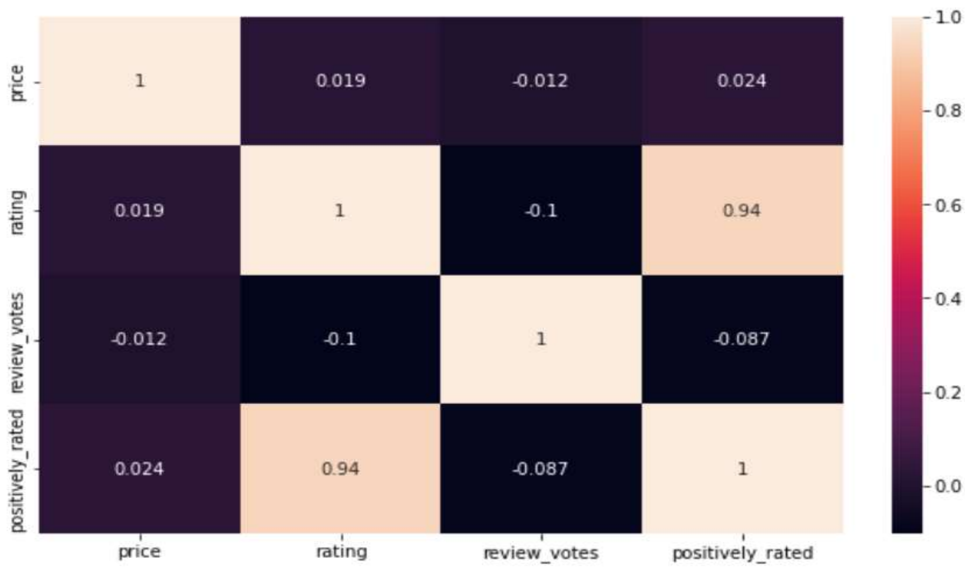


Fig. 3 Heat map of the respective variables

As we can see the positive_rated increases, recommendation of cell phones increases followed by the price variabl e. This is known as positive correlation. The variable review_votes has the highly negatively correlated. There is a condition where neither positive correlation nor negative correlation in between the variables. This situation is kno wn as no correlation in between the variables. From this heat map it can be conclude that positive_rated, price and review_votes are important variables.
In the positively rated variable there are 23606 reviews that shows for not recommendation for buying the cell pho nes while 11059 reviews showing the positive feedback about the cell phones. According to the figure 4, 31.9% re views positive recommendation for cell phones while 68.1% reviews not recommended for buying cell phones. Th e length of positive word is 293399 and the length of negative word is 204620 in attribute reviews.

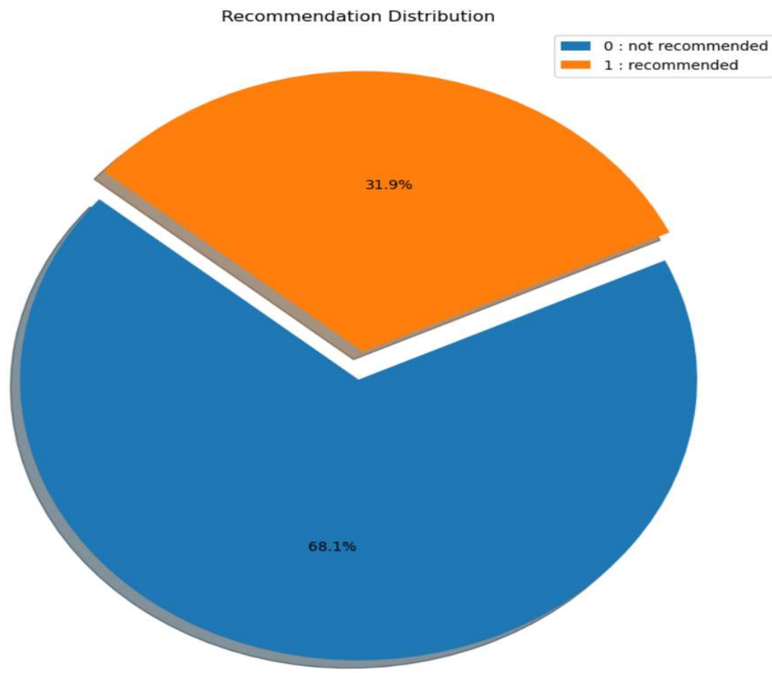


Fig.4 Reviews of recommendation

A supervised ML algorithm was used to classify the data and discover patterns in the dataset.
The scores that was estimated in the table 3, the F1_score which evaluate the model's accuracy through combines the precision and recall scores of a model in machine learning [36]. This accuracy measurements processes that how often a model right forecast across the whole dataset. The concept that F1_score uses that it is better to compute harmonic mean than simple mean or arithmatic mean. The higer F1_score represent the better quality of the classifiers. It is calculated by the formula:

$$\text{F1\_score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

In the table 3, the highest F1_score for SA is gain by the classifier AdaBoost_TFIDF which is almost 94.5%.
The Recall_score also known as sentivity analysis which is the positive events that we predicted correctly [37]. The recall or senstivity is given by the following formula:

$$\text{Recall\_score} = \frac{True\ positive}{True\ positive + False\ negative}$$

The Recall score of Sentiment analysis from table 3, the highest score is gain by the classifier NaiveBayes (Multi) _TFIDF which is 96%.

The area under the precision curve is known as average precision. The average precision selects a confidence from hard and subjective applications [38]. It tries to remove the dependency of selecting one confidence threshold value. It is measure by the following formula:

$$Average\ precision = \int_{r=0}^{1} p(r)dr$$

The highest average precision score is 98.2% which is gain by AdaBoost_TFIDF as well as LogReg_TFIDF.

Table 3 Performance metric measures

Model	F1_Score	Recall_Score	Average_Precision_Score
NaiveBayes(Multi)_Count	0.936	0.958	0.974
NaiveBayes(Berno)_Count	0.885	0.929	0.964
LogReg_Count	0.866	0.905	0.976
SVM_Count	0.937	0.921	0.977

Random Forest_Count	0.914	0.925	0.971
AdaBoost_Count	0.934	0.954	0.970
NaiveBayes(Multi)_TFIDF	0.934	0.960	0.981
NaiveBayes(Berno)_TFIDF	0.885	0.929	0.964
LogReg_TFIDF	0.906	0.858	0.982
SVM_TFIDF	0.905	0.854	0.974
AdaBoost_TFIDF	0.945	0.953	0.982

Figure 5 is the graphical representation of the classifiers.

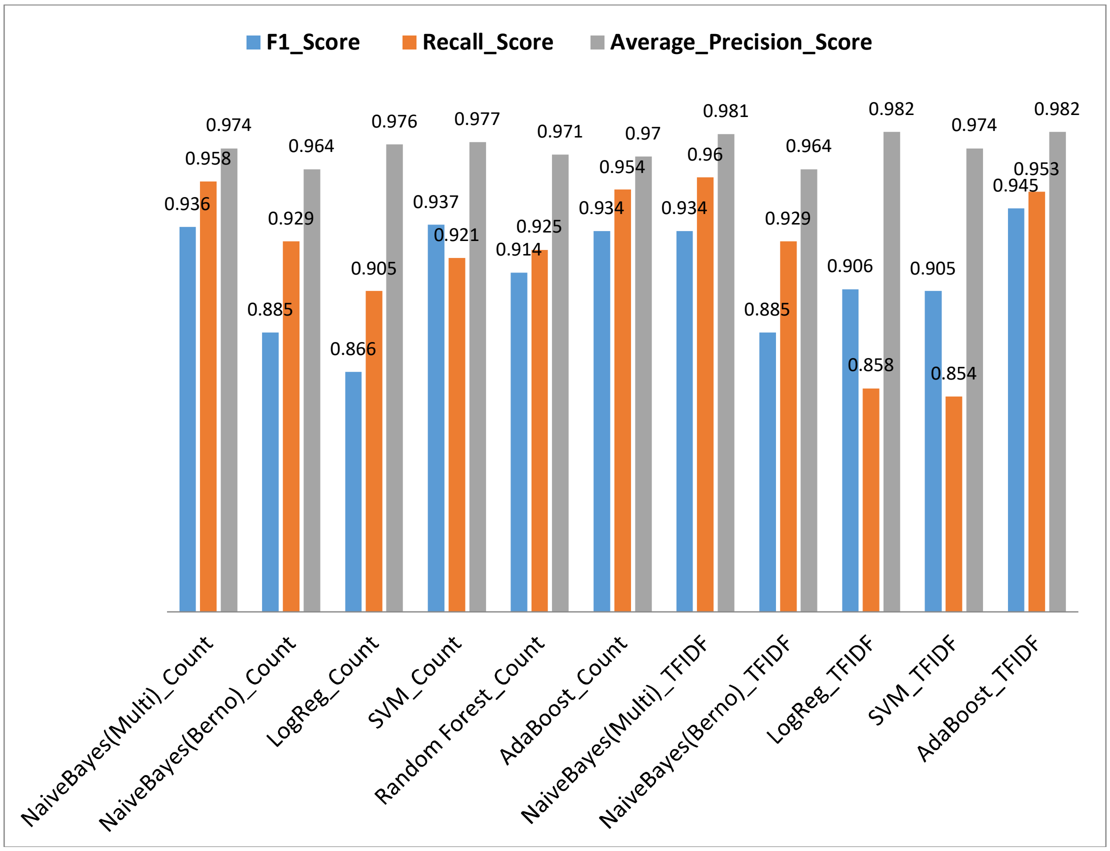


Fig.5 Performance representation of the classifiers

6. Conclusions

This study aimed to develop a model for the accurate prediction of sentiment in customer reviews. Data on customer's reviews on cell phones were collected from Kaggle. The proposed model classified positive, negative, and neutral sentiments. The dataset was preprocessed in several steps using different methods to train and improve ML algorithms to reduce dimensionality. The F1_score, Recall_score and Average_precision_score for sentiments of reviews were analyzed, as shown in Table 3 and figure 5. However, the results showed that the highest F1_score obtained by AdaBoost_TFIDF which is 94.5%, the highest Recall_score is 96% by NaiveBayes (Multi) _TFIDF as well as the highest Average_Precision_Score is 98.2% which is tie in between AdaBoost_TFIDF and LogReg_TFIDF. Keeping in mind the customer reviews in terms of sentiment analysis, it can be said that AdaBoost, NaiveBayes and logistic regression TFIDF are the better classifiers.

By improving the method used to identify sentiments in customer reviews, this study adds to the empirical literature. This study's findings have the potential to enhance the customer experience. The model's performance was enhanced by this study's class balancing methods. We used multiclass classification to conduct unbalanced sampling based on multiple methods, in contrast to the majority of previous studies that used binary classification. In future work, we will expand the current concentrate by applying deep learning models. A hyperparameter tuning model, for instance, could be used to enhance the current study's findings.

Studies have been carried out recently applying scientometric analysis to determine the growth of research production.
Aydin (2017) conducted the research on "Research Performance of Higher Education Institutions", the article intends to raise awareness of "research performance," which plays a crucial role in university competition. The study makes an effort to summarize the findings of a thorough literature evaluation in the area of higher education research performance in order to achieve this goal. First, basic literature on research performance is discussed together with its concept definition and indicators. Then, a thorough presentation of the variables affecting research performance followed. The study concludes with the provision of a conceptual framework that will be useful to all university staff.

References

[1] Doherty, N. F., & Ellis-Chadwick, F. (2010). Internet retailing: the past, the present and the future. International Journal of Retail & Distribution Management, 38(11/12), 943-965.

[2] Malbon, J. (2013). Taking fake online consumer reviews seriously. Journal of Consumer Policy, 36, 139-157.

[3] Gefen, D., Fresneda, J. E., & Larsen, K. R. (2020). Trust and distrust as artifacts of language: a latent semantic approach to studying their linguistic correlates. Frontiers in psychology, 11, 488398.

[4] Marucheck, A., Greis, N., Mena, C., & Cai, L. (2011). Product safety and security in the global supply chain: Issues, challenges and research opportunities. Journal of operations management, 29(7-8), 707-720.

[5] Mangold, W. G., & Smith, K. T. (2012). Selling to Millennials with online reviews. Business Horizons, 55(2), 141-153.

[6] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big data, 2, 1-14.

[7] Telser, L. G. (1960). Why should manufacturers want fair trade?. The journal of law and economics, 3, 86-105.

[8] Stoop, I. A., Billiet, J., Koch, A., & Fitzgerald, R. (2010). Improving survey response: Lessons learned from the European Social Survey. John Wiley & Sons.

[9] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.

[10] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-based systems, 89, 14-46.

[11] Sari, P. K., Alamsyah, A., & Wibowo, S. (2018). Measuring e-Commerce service quality from online customer review using sentiment analysis. Journal of Physics: Conference Series, 971(1), 12053. https://doi.org/10.1088/1742- 6596/971/1/012053

[12] Hariawansah, M. (2018). Analisis Sentimen ECommerce Dengan Menggunakan Metode Naive Bayes. In Department of Computer Science, Binus University Graduate Program.

[13] Noor, A., & Islam, M. (2019). Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–6. https://doi.org/10.1109/ICCCNT45670.201 9.8944436

[14] M. Raheem, M. Marong, N. K. Batcha, and R. Mafas, "Sentiment Analysis in E-Commerce: A Review on The Techniques and Algorithms," J. Appl. Technol. Innov., vol. 4, no. 1, p. 6, 2020.

[15] Sharma, M. (2020). Sentiment Analysis: An Introduction to Naive Bayes Algorithm. https://towardsdatascience.com/sentimentanalysis-introduction-to-naive-bayesalgorithm 96831d77ac91

[16] Handayani, Y., & Hakim, A. R. (2020). Sentiment Analysis of Bank BNI User Comments Using the Support Vector Machine Method. 2020 International Seminar on Application for Technology of Information and Communication (ISemantic), 202–207. https://ieeexplore.ieee.org/abstract/document/9234230/

[17] Marong, M., Batcha, N. K., & Mafas, R. (2020). Sentiment Analysis in E-Commerce: A Review on The Techniques and Algorithms. Journal of Applied Technology and Innovation (e-ISSN: 2600-7304), 4(1), 6. https://jati.sites.apiit.edu.my/files/2020/01/ Sentiment-Analysis-in-E-Commerce-AReview-on-The-Techniques-andAlgorithms.pdf

[18] T. Willianto, Supryadi, and A. W. (2020). Sentiment Analysis on E-Commerce Product using Machine Learning an Combination of TF-IDF and Backward Elimination. International Journal of Recent Technology and Engineering (IJRTE),. https://doi.org/10.35940/ijrte.F7889.038620

[19] Naresh A. Recommender system for sentiment analysis using machine learning models. Turk J Comput Math Educ. 2021;12(10):583–588.

[20] Karthik RV, Ganapathy S. A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. Appl Soft Comput. 2021;108:107396. doi:10.1016/j.asoc.2021. 107396

[21] Ren, S., Zhang, Y., Liu, Y., Sakao, T., Huisingh, D., & Almeida, C. M. (2019). A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions. Journal of cleaner production, 210, 1343-1365.

[22] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, 117-126.

[23] Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. International Journal on Advances in Software, 10(1), 1-20.

[24] Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data processing and text mining technologies on electronic medical records: a review. Journal of healthcare engineering, 2018(1), 4302425.

[25] Atenstaedt, R. (2017). Word cloud analysis of the BJGP: 5 years on. British Journal of General Practice, 67(658), 231-232.

[26] Maurya, R. K., Kumar, N., Srivastava, V., Kumar, S., & Kumar, V. (2022). Prognostic And Diagnostic Prediction Models Of Invasive Cancer In Women And Its Detection Accuracy Via Developing Machine Learning Techniques. Journal of Pharmaceutical Negative Results, 4282-4294.

[27] Polyakov, E. V., Voskov, L. S., Abramov, P. S., & Polyakov, S. V. (2020). Generalized approach to sentiment analysis of short text messages in natural language processing. Информационно-управляющие системы, (1 (104)), 2-14.

[28] Al-Obaydy, W. I., Hashim, H. A., Najm, Y. A., & Jalal, A. A. (2022). Document classification using term frequency-inverse document frequency and K-means clustering. Indonesian Journal of Electrical Engineering and Computer Science, 27(3), 1517-1524.

[29] Muchlinski, D., Siroky, D., He, J., & Kocher, M. (2016). Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. Political Analysis, 24(1), 87-103.

[30] Jaiswal, A., & Kumar, R. (2019). Review of machine learning algorithm on cancer classification for cancer prediction and detection. Int J Anal Exper Modal Anal (IJAEMA), 11(12), 1456-1466.

[31] Somanth, R., Chowdary, R. T., Akhil, B. J., Sanjeet1d, R., & Harsha, S. S. (2024). BREAST CANCER PREDICTION USING MACHINE LEARNING. BREAST CANCER, 18(01).

[32] Huang, G., Cheng, A., & Gao, Y. (2022, October). Machine Learning improvements to the accuracy of predicting Specific Language Impairment. In 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 553-566). IEEE.

[33] Chaurasia, V., & Pal, S. (2022). Ensemble technique to predict breast cancer on multiple datasets. The Computer Journal, 65(10), 2730-2740.

[34] https://www.kaggle.com/code/srinivasav22/nlp-amazon-unlocked-mobile/input. Accessed on 12th June 2024.

[35] Chaurasia, V., & Pal, S. (2022). An ensemble framework-stacking and feature selection technique for detection of breast cancer. International Journal of Medical Engineering and Informatics, 14(3), 240-251.

[36] Yacouby, R., & Axman, D. (2020, November). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the first workshop on evaluation and comparison of NLP systems (pp. 79-91).

[37] Mahmud, M. H. (2020). The applications of data mining and machine learning in Bangladesh for disease pattern analysis and prediction (Doctoral dissertation, Brac University).

[38] Méndez Méndez, A. E., Cartwright, M., Bello, J. P., & Nov, O. (2022). Eliciting confidence for improving crowdsourced audio annotations. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1-25.