FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 1, January, 2025, pp 46 - 55
DOI: https://doi.org/10.33003/fjs-2025-0901-2795

X (FORMALLY TWITTER) PRODUCT CLASSIFICATION USING NAÏVE BASE

***[1]Orie, C. E., [1]Egwali, A. O. and [2]Amadin, F. I.**

[1]Department of Computer Science, Benson Idahosa University, Benin City, Nigeria
[2]Department of Computer Science, University of Benin, Benin City, Nigeria

*Corresponding authors' email: corie@biu.edu.ng

ABSTRACT

Finding polarities in a textual data is very important in sentiment analysis. Navie Bayes is one of the most effective machine learning classifier techniques and a probabilistic classier that applies Bayes theorem and assuming feature independence for the classification of data. The objective of this research is to implement the Naïve Bayes algorithm for the classification of sentiment in the context of X (Formally Twitter) product classification. Our aim is developing a model that can conveniently classify product-related text into positive and negative sentiment categories. The process begins with the collection of customer product reviews from X (Formally Twitter) and vectorizing each reviews making a long array of unique words, whose attributes include the independent vectors while assigned values are the number of times each independent vector appear in the product review. We had a total of one hundred and thirty-three (133) unique words in the training set for both the positive and Negative statements of which ten (10) documents were with positive (+) outcomes and ten (10) documents were with negative (-) outcomes. From the product reviews we derived the probability of each positive outcomes of independent word, probability of each negative outcomes of independent words and then the probability of each word in the product review. Our results shows that the naïve bayes classifier is a good classification technique and will be effective for both large and small businesses in making decisions related to their product development, marketing campaigns and customer support.

Keywords: Polarities, Textual, Sentiment, Vectorizing, Unique

INTRODUCTION

Sentiment classification involves the analysis of people's opinions in textual forms and classifying these opinions as either positive, negative or neutral polarities depending on the scenarios in which they are used. The two major aspects in natural language processing are sentiment analysis and product classification which a lot of researchers have looked into. With the massive increase of online platform, the classification and analysis of sentiments towards customers feedback and the accurate classification of products into their different polarities have become greatly important enabling businesses to make beneficial decisions that will improve customer experiences. A lot of challenges with sentiment classification is recurrent, the major problem is classifying an individual's opinion or sentiments as a positive statement or a negative statement. A second challenge is that individuals don't continuously specify opinions within the same manner (Shukla and Mishra, 2016).

Several researchers have contributed to the understanding and improvement of sentiment analysis and product classification techniques, Pang and Lee (2008) provide a comprehensive overview of sentiment analysis, covering sentiment lexicons, machine learning techniques, and associated challenges. While Naïve Bayes is not specially explored in their work, in this research the Naïve Bayes machine learning techniques will be employed to analyze already classified product reviews from X (formally twitter) platform. This classification is useful to every individual, and business owners.it will help organizations track how their products is perceived outside, Governments and political organizations can also use this sentimental classification to gauge public opinion on various policies and election campaigns. Documents can be classified in three ways: the supervised, unsupervised and the semi-supervised methods Korde (2012). Laila *et al;* (2018) in their research choose the Naive Bayes technique to carry out sentiment analysis because of its high level of accuracy. They divided their research into two parts which include the before and the after endorsement. They collected data by a preprocessing method and the classification process was carried out using the confusion matrix. Their result proves that the Naive Bayes has an excellent accuracy rate more than 84%. Although, the negative sentiment rose by 12.51%.

Sentiment analysis of tweet data is significant to both individuals, students, schools, businessmen, politicians, organizations etc. Sentiment analysis can be an excellent source of information and can provide insights that can: Determine marketing strategy, improve campaign success, improve product messaging, improve customer service etc.

However, the study of sentiment Analysis, if done properly, is exceptionally complex and is actually a field of study, not just a future in social media tool (Bing, 2012).

There are several techniques for analysing sentiments of tweets like the supervised machine learning technique and the unsupervised machine learning technique but The supervised machine learning technique is the most used machine learning technique for sentiment analysis to achieve the best accuracy Umarani *et al;* (2021).

The aim of this research is to develop a Naïve Bayes-based sentiment analysis model for classifying tweets from X (formerly Twitter) into different polarities which include positive and negative categories.

MATERIALS AND METHODS

Sentiment Classification is a very important aspect in today's research and various researchers has taken time to carry out this study using different machine learning techniques. There are numerous machine learning classification algorithms and Naïve Bayes is one of the best as it gives accurate results even in the presence of an elaborate dataset Vidhya and Aghila (2010).

Pang *et al*; (2002) in a conference proceedings analyzed the classification of various documents not only by their topics but also by the total sentiments, such as ascertaining the

probability of an opinion to be either a positive sentiment or a negative sentiment. The dataset comprises of movie reviews and they noted that machine learning techniques performs better than human standard. In concluding their study, they examined other factors that makes classifying peoples sentiments classification a challenge.

Suchdev *et al;* (2014) analyses individuals' opinions in sentiments regarding some organization. In their work the computation of some basic sentimental scoring was reviewed which they classified as positive or negative this was done to assist companies by helping them with sentiments regarding their products from various individuals. During their research, they used the Sanders analytic dataset in analyzing collected tweets. The first stage is their work was pre-processing the data, the hybrid approach was applied and 100% accuracy was achieved.

In a work by Singla *et al*; (2017), over 4,000,00 sentiments were analyzed and the classification was shared into positive sentiment and negative sentiments using machine learning classification algorithm. Among the numerous machine learning techniques, Naïve Bayesians algorithm, the SVM algorithm and the Decision - Tree were used for analyzing the collected sentiments.

Dey *et al;* (2016) project was to check the sentiment content of some movie reviews and hotel reviews with the analysis of both. In their work, A statistical classification approach was used to capture elements of the polarity in a sentence alongside the subjectivity of the sentence. They also talked on two major supervised techniques which are K Nearest Neighbor (KNN) and Naïve Bayes' a comparison of both was carried out. At the end of their work, the Naïve bayes gave better result for movie reviews compared to the K-NN although for hotel reviews both algorithms resulted in almost similar accuracies.

Agarwal *et al*; (2011) examined sentimental classification on Twitter data by using the state-of-the-art unigram model as measuring point, they derived a total estimate of 4% in the two tasks which was classified.

Fang and Zhan (2015) in their paper, tackled a challenge with sentiment analysis which is categorizing sentiment polarity which is a major issue with sentiment analysis. Their dataset was collected from Amazon.com on product reviews. The analysis for both sentence-level classification and review-level classification were done giving excellent performance.

Narayanan *et al*; (2013) have investigated various approaches that can improve the result of the Naïve Bayes algorithm for classifying people's sentiments. In this work, joining different effective approaches which include reverse handling, words ngrams and feature-selection by shared data resulted at good accuracy. An 88.80% of accuracy on the popular IMDB movie reviews dataset was achieved at the end.

Hemalatha *et al*., (2014) in their paper presented a system which collects Tweets from social networking sites, they used the machine learning techniques to analyze tweets and achieved accurate prediction for business organizations.

Classification of Text

Classification of text is the process of classifying reviews as either positive or negative on the basis of their sentiment Singla (2017). Text Classification is also called:

i. Text Categorization
ii. Document Classification
iii. Document Categorization

There are two methods of classifying text they are:

i. Manual classification
ii. Automatic classification.

In this work, both methods of classification will be used using a training sample of already classified examples sourced from twitter.

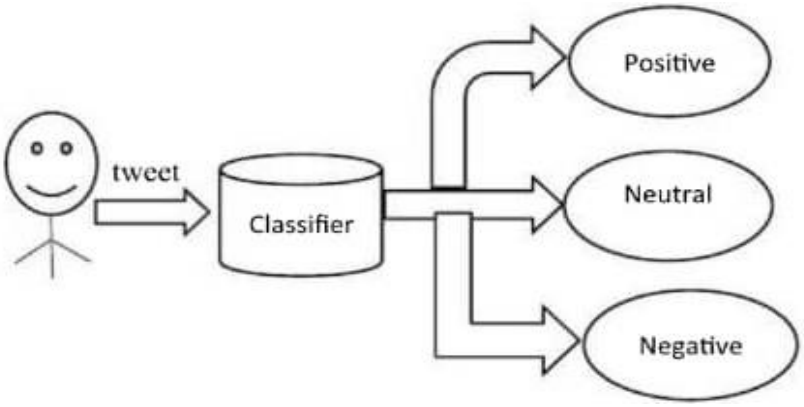

Figure 1: Text classification

Naïve Bayes Machine Learning Technique

Naïve Bay's is among the most effective algorithm, it is referred to as a classification technique based on Bayes' Theorem and works with an assumption of prediction. The Naive Bayes (NB) classifier is the most frequently used method for classifying text documents, it assumes that the probabilities being combined are independent of each other Rai and Jaiswal (2017).

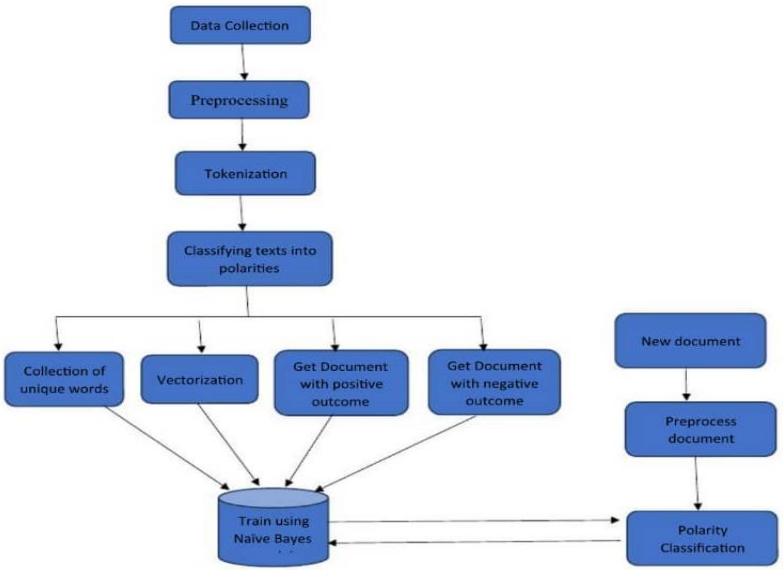

Figure 2: Proposed methodology

Data Collection

The document dataset comprises of sentimental opinions or reviews from various individuals regarding an entity. This entity can either be topics of interest, or any product. A proper data collection is very effective to form a document dataset. Data collection is the first step in sentiment analysis and the selection of data is very crucial for an effective analysis. A total of twenty data relating to product review were collected from twitter. This will also form our training set. The following are the collected positive and negative tweets from twitter on product reviews:

Doc	Text	Class
1	So there is no way for me to plug it in here in the US unless I go by a con-verter	0
2.	Good case Excellent Value.	1
3.	Great for the Jawbone.	1
4.	Tied to charger for Conversations lasting more than 45 Minutes. MAJOR PROBLEMS!!	0
5.	(The mic is Great)	1
6.	I have to jiggle the plug to get it to line up right to get Decent Volume.	0
7.	If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one.	0
8.	If you are Razr owner...you must have this!	1
9.	Needless to say, I wasted my money.	0
10.	What a waste of money and time!	0
11.	And the sound quality is great.	1
12.	He was very impressed when going from the original battery to the extended battery.	1
13.	If the two were separated by a mere 5+ ft I started to notice excessive static and garbled sound from the headset.	0
14.	Very good quality though	1
15.	The design is very odd, as the ear "clip" is not very comfortable at all.	0
16.	Highly recommend for anyone who has a blue tooth phone.	1
17.	I advise EVERYONE DO NOT BE FOOLED!	0
18.	So Far So Good!.	1
19.	Works Great!.	1
20.	It clicks into place in a way that makes you wonder how long that mechanism would last.	0

Figure 3: Data set

Preprocessing

The notable reason for preprocessing during classification of sentiments or opinions is limiting the number of features tweet and also to reduce the complex nature with the sentiment classification Mohan *et al,* (2016). Since the data has been collected, they are then preprocessed *which is done to remove all irrelevant details like punctuation marks that lacks dictionary meaning Orie et al; (2024).*

Tokenization

Tokenization refers to the act of splitting a tweet or sentence into independent vectors that serve as input for various natural language processing algorithms. This involves splitting the text into words and then discarding the nonrelevant words Kanade *et al*; (2019). Basically, there are different processing stages for tokenization and they include: removing of stop words and punctuation marks. Let's take an example of a tokenization process which splits a tweet into independent vector of words by removing irrelevant details like punctuation mark.

Here is one of the collected data sets:

Highly recommend for anyone who has a blue tooth phone.
Example of Tokenization:

Highly, recommend, for, anyone, who, has, a, blue, tooth, phone.

Classifying Texts into Polarities

Documents can either be positive (+) or negative (-) depending on the reviewer's sentiment. Classifying texts is important for the analysis of tweet. To classify text into different polarities the following steps needs to be considered:

i. Training sets should be collected from any of the social media platforms like newspapers, facebuk, and used for the analysis.
ii. Each document has to be represented by vector of words. These independent words known as unique words will be counted.
iii. Transform the tweet to feature-sets where the attributes include all independent vectors and the values include the specific times a vector occur in the collected tweet.

The training sets should be used to estimate:
(a) P (+)
(b) P (-)
(c) P (doc/+)
(d) P (doc/-)

Collection of Unique Words

Unique words are the independent vector of words present in the bag of words collected. This is important to accurately calculate the probability of each word present in a statement depending on the number of times they were used.

There are one hundred and thirty-three (133) unique words in the training set for both the positive and Negative statements. The unique words are: so, there, is, no, way, for, me, to, plug, it, in, here, the, US, unless, I, go, by, a, converter, good, case, excellent, value, great, jawbone, tied, charger, conversation, lasting, more, than, 45, minutes, major, problem, mic, have, jiggle, get, line, up, right, decent, volume, if, you, several, dozen, or, hundred, contracts, then, imagine, fun, sending, each, them, one, are, Razr, owner, must, this, needless, say, wasted, my, money, what, waste, of, and, time, sound, quality, he, was, very, impressed, when, going, from, original, battery, extended, two, were, separated, mere, 5, ft, started, notice, excessive, static, garbled, headset, though, design, odd, as, ear, clip, not, comfortable, at, all, highly, recommended, anyone, who, has, Bluetooth, phone, advise, everyone, do, be, fooled, far, works, clicks, into, place, that, makes, wonder, how, long, mechanism, would, last. These unique words will further be classified into two polarities (positive and negative) based on their classification.

Vectorization

This is done to assign numerical values to the individual words depending on the number of times they appear in the dataset. Vectorization is crucial as the Naïve Bayes Machine learning model can only understand numerical information since the model is a mathematical model Shah (2021). Below is a representation of our vectorized dataset.

Doc	so	there	is	no	way	for	me	to	plug	it	in	here	the	US	Un-less	i	go	by	a	Con-verter	good	case	exce llent	Va-lue	great
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
3	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0		0	0	1
6	0	0	0	0	0	0	0	4	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	1	0	1	1	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
15	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
18	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Figure 4: Vectorization of collected data

Documents with positive (+) outcomes

There are a total of ten (10) documents with positive (+) outcomes in the collected vocabulary and they include:

2. Good case, Excellent value.
3. Great for the jawbone.
5. The mic is great*.
8. If you are Razr owner...you must have this!
11. And the sound quality is great.
12. He was very impressed when going from the original battery to the extended batter
14. Very good quality though
16. Highly recommend for anyone who has a blue tooth phone.
18. So Far So Good!
19. Works great

Doc	so	there	is	no	way	for	me	to	plug	it	in	here	the	US	Un-less	i	go	by	a	Con-verter	good	case	Exce-llent	value	great	Jaw-bone
1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
4	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	4	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	1	0	1	1	0	0	0	0	0	0	0
15	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0		0	0	1	0	0	0	0	0	0	0

Figure 5: Representation of Document with positive outcome

Documents with negative (-) outcomes

There are also ten documents with negative (-) outcomes.

1. So there is no way for me to plug it in here in the US unless I go by a converter.
4. Tied to charger for conversations lasting more than 45 minutes. MAJOR PROBLEMS!!
6. I have to jiggle the plug to get it to line up right to get decent volume.
7. If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one.
9. Needless to say, I wasted my money.
10. What a waste of money and time!
13. If the two were separated by a mere 5+ ft I started to notice excessive static and garbled sound from the headset.
15. The design is very odd, as the ear "clip" is not very comfortable at all
17. I advise EVERYONE DO NOT BE FOOLED!
20. It clicks into place in a way that makes you wonder how long that mechanism would last.

Doc	so	there	is	no	way	for	me	to	plug	it	in	here	the	US	Un-less	i	go	by	a	Con-verter	good	case	Exce llent	value	great	Jaw bone	tied
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
5	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
16	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Figure 6: Representation of Documents with Negative outcomes

Train the Naïve Bayes Classifier

Naïve Bayes classifier is a very good classifier model where the implementation of model is to train a system and classify it to the corresponding category Vidhya and Aghila (2010) making the naïve bayes classifier easier than other machine learning techniques. The machine learning Classification algorithm was named after Thomas Bayes in the year 1702 to 1761 Hemalatha *et al*; (2014). In Calculating the class prior: Estimate the prior probability of each class based on the training data. This involves counting the number of all document in every class and dividing them by all available documents. We do so for both the positive outcomes and the negative outcomes to achieve both Naïve Bayes probabilities.

Probability of Positive (+) Outcomes

Using the independent vector of words present in the document with positive outcomes, the Computation of all the positive polarities in the training set has to be derived for accurate analysis that is, the P(good/+), P(case/+), p(excellent/+), P(value/+), P(for/+), P(the/+), P(great/+), P(jawbone/+), P(is/+), P(mic/+), P(have/+), P(if/+), P(you/+), P(are/+), P(Razr/+), P(owner/+), P(must/+), P(this/+),

P(and/+), P(sound/+), P(quality/+), P(to/+), P(he/+), P(was/+), P(very/+), P(impressed/+), P(when/+), P(going/+), P(form/+), P(original/+), P(battery/+), P(extended/+), P(though/+), P(a/+), P(highly/+), P(recommended/+), P(anyone/+), P(who/+), P(has/+), P(bluetooth/+), P(phone/+), P(so/+), P(far/+), P(works/+).

$P\ (Document/-/) = \frac{10}{20} = 0.5$

Where 10 is the total No. of positive document and 20 is the No. of collected document on product review from X (formally twitter).

There is a total of 133 unique words in the dataset. To calculate the probability of positive outcomes, Let m be the number of words in the positive (+) case which is a total of 57 and m_j the number of times the word j occurs in this positive (+) cases

Therefore:

$$P\ (wj/+/) = \frac{m_j + 1}{m + /vocabulary/}$$

P(good/+) = $\frac{3+1}{57+133}$ = 0.0211
P(excellent/+) = $\frac{1+1}{57+133}$ = 0.0105
P(for/+) = $\frac{2+1}{57+133}$ = 0.0158
P(great/+) = $\frac{4+1}{57+133}$ = 0.0263
P(is/+) = $\frac{2+1}{57+133}$ = 0.0158
P(have/+) = $\frac{1+1}{57+133}$ = 0.0105
P(you/+) = $\frac{1+1}{57+133}$ = 0.0105
P(Razr/+) = $\frac{1+1}{57+133}$ = 0.0105
P(must/+) = $\frac{1+1}{57+133}$ = 0.0105
P(and/+) = $\frac{1+1}{57+133}$ = 0.0105
P(quality/+) = $\frac{2+1}{57+133}$ = 0.0158
P(he/+) = $\frac{1+1}{57+133}$ =0.0105
P(very/+) = $\frac{2+1}{57+133}$ =0.0158
P(when/+) = $\frac{1+1}{57+133}$ =0.0105
P(from/+) = $\frac{1+1}{57+133}$ =0.0105
P(battery/+) = $\frac{1+1}{57+133}$ =0.0105
P(though/+) = $\frac{1+1}{57+133}$ =0.0105
P(highly/+) = $\frac{1+1}{57+133}$ =0.0105
P(anyone/+) = $\frac{1+1}{57+133}$ =0.0105
P(has/+) = $\frac{1+1}{57+133}$ =0.0105
P(phone/+) = $\frac{1+1}{57+133}$ =0.0105
P(far/+) = $\frac{1+1}{57+133}$ = 0.0105
P(case/+) = $\frac{1+1}{57+133}$ = 0.0105
P(value/+) = $\frac{1+1}{57+133}$ = 0.0105
P(the/+) = $\frac{4+1}{57+133}$ = 0.0263
P(jawbone/+) = $\frac{1+1}{57+133}$ = 0.0105
P(mic/+) = $\frac{1+1}{57+133}$ = 0.0105
P(if/+) = $\frac{1+1}{57+133}$ = 0.0105
P(are/+) = $\frac{1+1}{57+133}$ = 0.0105
P(owner/+) = $\frac{1+1}{57+133}$ = 0.0105
P(this/+) = $\frac{1+1}{57+133}$ =0.0105
P(sound/+) = $\frac{1+1}{57+133}$ =0.0105
P(to/+) = $\frac{1+1}{57+133}$ = 0.0105
P(was/+) = $\frac{1+1}{57+133}$ = 0.0105
P(impressed/+) = $\frac{1+1}{57+133}$ = 0.0105
P(going/+) = $\frac{1+1}{57+133}$ = 0.0105
P(original/+) = $\frac{1+1}{57+133}$ = 0.0105
P(extended/+) = $\frac{1+1}{57+133}$ =0.0105
P(a/+) = $\frac{1+1}{57+133}$ = 0.0105
P(recommended/+) = $\frac{1+1}{57+133}$ = 0.0105
P(who/+) = $\frac{1+1}{57+133}$ = 0.0105
P(Bluetooth/+) = $\frac{1+1}{57+133}$ = 0.0105
P(so/+) = $\frac{2+1}{57+133}$ = 0.0158
P(works/+) = $\frac{1+1}{57+133}$ = 0.0105

Documents with negative (-) outcomes

There are also ten documents with negative (-) outcomes from the collected tweets on product review from X (formally twitter). The documents with positive outcomes include:

1 So there is no way for me to plug it in here in the US unless I go by a converter.

4. Tied to charger for conversations lasting more than 45 minutes. MAJOR PROBLEMS!!

6. I have to jiggle the plug to get it to line up right to get decent volume.

7. If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one.

9. Needless to say, I wasted my money.

10. What a waste of money and time!

13. If the two were separated by a mere 5+ ft I started to notice excessive static and garbled sound from the headset.

15. The design is very odd, as the ear "clip" is not very comfortable at all

17. I advise EVERYONE DO NOT BE FOOLED!

20. It clicks into place in a way that makes you wonder how long that mechanism would last.

Probability of Negative (-) Outcomes

The probability of the negative outcome present in the collected tweet needs to be derived just like it was derieved for the positive outcomes. i.e the P(so/-), P(there/-), P(is/-), P(no/-), P(way/-), P(for/-), P(me/-), P(to/-), P(plug/-), P(it/-), P(in/-), P(the/-), P(US/-), P(unless/-), P(i/-), P(go/-), P(by/-), P(a/-), P(covnerter/-), P(tied/-), P(charger/-),P(conversation/-),P(lasting/-),P(more/-),P(than/-),P(45/-),P(minutes/-),P(major/-),P(problems/-), P(have/-), P(jiggle/-), P(get/-), P(line/-), P(up/-), P(right/-), P(decent/-), P(volume/-), P(if/-), P(you/-), P(several/-), P(dozen/-), P(or/-), P(hundred/-), P(contracts/-), P(then/-), P(imagine/-), P(fun/-), P(sending/-), P(each/-), P(them/-), P(one/-), P(needless/-), P(say/-), P(wasted/), P(my/-), P(money/-), P(what/-), P(waste/-), P(of/-), P(and/-), P(time/-), P(sound/-), P(from/-), P(two/-), P(were/-), P(separated/-), P(mere/-), P(5/-), P(ft/-), P(started/-), P(notice/-), P(excessive/-), P(static/-), P(garbled/-), P(headset/-), P(very/-), P(odd/-), P(as/-), P(ear/-), P(clip/-), P(not/-), P(comfortable/-), P(at/-), P(all/-), P(advice/-), P(everyone/-), P(do/-), P(be/-), P(fooled/-), P(clicks/-), P(into/-), P(place/-), P(that/-), P(makes/-), P(wonder/-), P(how/-), P(long/-), P(mechanism/-), P(would/-), P(last/-).

$P\ (Document/-/) = \frac{10}{20} = 0.5$

Let m be the number of words in the negative (-) case which is a total of 141 and m_j the number of times the word j occurs in this negative (-) cases

Let:

$$P\ (wj/-/) = \frac{m_j + 1}{m + /vocabulary/}$$

P(so/-) = 1 + 1 / 141 + 133 = 0.0073
P(is/-) = 3 + 1 / 141 + 133 = 0.0146
P(way/-) = 1 + 1 / 141 + 133 = 0.0073
P(me/-) = 1 + 1 / 141 + 133 = 0.0073
P(plug/-) = 2 + 1 / 141 + 133 = 0.0109
P(in/-) = 3 + 1 / 141 + 133 = 0.0146
P(US/-) = 1 + 1 / 141 + 133 = 0.0073
P (1/-) = 5 + 1 / 141 + 133 = 0.0219
P(by/-) = 3 + 1 / 141 + 133 = 0.0146
P(converter/-) = 1 + 1 / 141 + 133 = 0.0073
P(charger/-) = 1 + 1 / 141 + 133 = 0.0073
P(lasting/-) = 1 + 1 / 141 + 133 = 0.0073
P(than/-) = 1 + 1 / 141 + 133 = 0.0073
P(minutes/-) = 1 + 1 / 141 + 133 = 0.0073
P(problems/-) = 1 + 1 / 141 + 133 = 0.0073
P(jiggle/-) = 1 + 1 / 141 + 133 = 0.0073
P(line/-) = 1 + 1 / 141 + 133 = 0.0073
P(right/-) = 1 + 1 / 141 + 133 = 0.0073
P(volume/-) = 1 + 1 / 141 + 133 = 0.0073
P(you/-) = 2 + 1 / 141 + 133 = 0.0109
P(dozen/-) = 1 + 1 / 141 + 133 = 0.0073
P(hundred/-) = 1 + 1 / 141 + 133 = 0.0073
P(then/-) = 1 + 1 / 141 + 133 = 0.0073
P(fun/-) = 1 + 1 / 141 + 133 = 0.0073

P(there/-) = 1 + 1 / 141 + 133 = 0.0073
P(no/-) = 1 + 1 / 141 + 133 = 0.0073
P(for/-) = 2 + 1 / 141 + 133 = 0.0109
P(to/-) = 8 + 1 / 141 + 133 = 0.0328
P(it/-) = 3 + 1 / 141 + 133 = 0.0146
P(the/-) = 7 + 1 / 141 + 133 = 0.0292
P(unless/-) = 1 + 1 / 141 + 133 = 0.0073
P(go/-) = 1 + 1 / 141 + 133 = 0.0073
P(a/-) = 4 + 1 / 141 + 133 = 0.0182
P(tied/-) = 1 + 1 / 141 + 133 = 0.0073
P(conversation/-) = 1 + 1 / 141 + 133 = 0.0073
P(more/-) = 1 + 1 / 141 + 133 = 0.0073
P (45/-) = 1 + 1 / 141 + 133 = 0.0073
P(major/-) = 1 + 1 / 141 + 133 = 0.0073
P(have/-) = 2 + 1 / 141 + 133 = 0.0109
P(get/-) = 2 + 1 / 41 + 133 = 0.0109
P(up/-) = 1 + 1 / 141 + 133 = 0.0073
P(decent/-) = 1 + 1 / 141 + 133 = 0.0073
P(if/-) = 2 + 1 / 141 + 133 =0.0109
P(several/-) = 2 + 1 / 141 + 133 = 0.0109
P(or/-) = 1 + 1 / 141 + 133 = 0.0073
P(contracts/-) = 1 + 1 / 141 + 133 = 0.0073
P(imagine/-) = 1 + 1 / 141 + 133 = 0.0073
P(sending/-) = 1 + 1 / 141 + 133 = 0.0073

P(so/-) = 1 + 1 / 141 + 133 = 0.0073
P(is/-) = 3 + 1 / 141 + 133 = 0.0146
P(way/-) = 1 + 1 / 141 + 133 = 0.0073
P(me/-) = 1 + 1 / 141 + 133 = 0.0073
P(plug/-) = 2 + 1 / 141 + 133 = 0.0109
P(in/-) = 3 + 1 / 141 + 133 = 0.0146
P(US/-) = 1 + 1 / 141 + 133 = 0.0073
P (1/-) = 5 + 1 / 141 + 133 = 0.0219
P(by/-) = 3 + 1 / 141 + 133 = 0.0146
P(converter/-) = 1 + 1 / 141 + 133 = 0.0073
P(charger/-) = 1 + 1 / 141 + 133 = 0.0073
P(lasting/-) = 1 + 1 / 141 + 133 = 0.0073
P(than/-) = 1 + 1 / 141 + 133 = 0.0073
P(minutes/-) = 1 + 1 / 141 + 133 = 0.0073
P(problems/-) = 1 + 1 / 141 + 133 = 0.0073
P(jiggle/-) = 1 + 1 / 141 + 133 = 0.0073
P(line/-) = 1 + 1 / 141 + 133 = 0.0073
P(right/-) = 1 + 1 / 141 + 133 = 0.0073
P(volume/-) = 1 + 1 / 141 + 133 = 0.0073
P(you/-) = 2 + 1 / 141 + 133 = 0.0109
P(dozen/-) = 1 + 1 / 141 + 133 = 0.0073
P(hundred/-) = 1 + 1 / 141 + 133 = 0.0073
P(then/-) = 1 + 1 / 141 + 133 = 0.0073
P(fun/-) = 1 + 1 / 141 + 133 = 0.0073

P(there/-) = 1 + 1 / 141 + 133 = 0.0073
P(no/-) = 1 + 1 / 141 + 133 = 0.0073
P(for/-) = 2 + 1 / 141 + 133 = 0.0109
P(to/-) = 8 + 1 / 141 + 133 = 0.0328
P(it/-) = 3 + 1 / 141 + 133 = 0.0146
P(the/-) = 7 + 1 / 141 + 133 = 0.0292
P(unless/-) = 1 + 1 / 141 + 133 = 0.0073
P(go/-) = 1 + 1 / 141 + 133 = 0.0073
P(a/-) = 4 + 1 / 141 + 133 = 0.0182
P(tied/-) = 1 + 1 / 141 + 133 = 0.0073
P(conversation/-) = 1 + 1 / 141 + 133 = 0.0073
P(more/-) = 1 + 1 / 141 + 133 = 0.0073
P (45/-) = 1 + 1 / 141 + 133 = 0.0073
P(major/-) = 1 + 1 / 141 + 133 = 0.0073
P(have/-) = 2 + 1 / 141 + 133 = 0.0109
P(get/-) = 2 + 1 / 41 + 133 = 0.0109
P(up/-) = 1 + 1 / 141 + 133 = 0.0073
P(decent/-) = 1 + 1 / 141 + 133 = 0.0073
P(if/-) = 2 + 1 / 141 + 133 =0.0109
P(several/-) = 2 + 1 / 141 + 133 = 0.0109
P(or/-) = 1 + 1 / 141 + 133 = 0.0073
P(contracts/-) = 1 + 1 / 141 + 133 = 0.0073
P(imagine/-) = 1 + 1 / 141 + 133 = 0.0073
P(sending/-) = 1 + 1 / 141 + 133 = 0.0073

P(each/-) = 1 + 1 / 141 + 133 = 0.0073
P(one/-) = 2 + 1 / 141 + 133 = 0.0109
P(say/-) = 1 + 1 / 141 + 133 = 0.0073
P(my/-) = 1 + 1 / 141 + 133 = 0.0073
P(what/-) = 1 + 1 / 141 + 133 = 0.0073
P(of/-) = 1 + 1 / 141 + 133 = 0.0073
P(time/-) = 1 + 1 / 141 + 133 = 0.0073
P(from/-) = 1 + 1 / 141 + 133 = 0.0073
P(were/-) = 1 + 1 / 141 + 133 = 0.0073
P(mere/-) = 1 + 1 / 141 + 133 = 0.0073
P(ft/-) = 1 + 1 / 141 + 133 = 0.0073
P(notice/-) = 1 + 1 / 141 + 133 = 0.0073
P(static/-) = 1 + 1 / 141 + 133 = 0.0073
P(headset/-) = 1 + 1 / 141 + 133 = 0.0073
P(odd/-) = 1 + 1 / 141 + 133 = 0.0073
P(ear/-) = 1 + 1 / 141 + 133 = 0.0073
P(not/-) = 1 + 1 / 141 + 133 = 0.0073
P(at/-) = 1 + 1 / 141 + 133 = 0.0073
P(advise/-) = 1 + 1 / 141 + 133 = 0.0073
P(do/-) = 1 + 1 / 141 + 133 = 0.0073
P(fooled/-) = 1 + 1 / 141 + 133 = 0.0073
P(into/-) = 1 + 1 / 141 + 133 = 0.0073
P(that/-) = 1 + 1 / 141 + 133 = 0.0073
P(wonder/-z) = 1 + 1 / 141 + 133 = 0.0073
P(long/-) = 1 + 1 / 141 + 133 = 0.0073
P(would/-) = 1 + 1 / 141 + 133 = 0.0073

P(them/-) = 1 + 1 / 141 + 133 = 0.0073
P(needless/-) = 1 + 1 / 141 + 133 = 0.0073
P(wasted/-) = 1 + 1 / 141 + 133 = 0.0073
P(money/-) = 2 + 1 / 141 + 133 = 0.0109
P(waste/-) = 1 + 1 / 141 + 133 = 0.0073
P(and/-) = 2 + 1 / 141 + 133 = 0.0109
P(sound/-) = 1 + 1 / 141 + 133 = 0.0073
P(two/-) = 1 + 1 / 141 + 133 = 0.0073
P(separated/-) = 1 + 1 / 141 + 133 = 0.0073
P(5/-) = 1 + 1 / 141 + 133 = 0.0073
P(started/-) = 1 + 1 / 141 + 133 = 0.0073
P(excessive/-) = 1 + 1 / 141 + 133 = 0.0073
P(garbled/-) = 1 + 1 / 141 + 133 = 0.0073
P(very/-) = 2 + 1 / 141 + 133 = 0.0109
P(as/-) = 1 + 1 / 141 + 133 = 0.0073
P(clip/-) = 1 + 1 / 141 + 133 = 0.0073
P(comfortable/-) = 1 + 1 / 141 + 133 = 0.0073
P(all/-) = 1 + 1 / 141 + 133 = 0.0073
P(everyone/-) = 1 + 1 / 141 + 133 = 0.0073
P(be/-) = 1 + 1 / 141 + 133 = 0.0073
P(clicks/-) = 1 + 1 / 141 + 133 = 0.0073
P(place/-) = 1 + 1 / 141 + 133 = 0.0073
P(makes/-) = 1 + 1 / 141 + 133 = 0.0073
P(how/-) = 1 + 1 / 141 + 133 = 0.0073
P(mechanism/-) = 1 + 1 / 141 + 133 = 0.0073
P(last/-) = 1 + 1 / 141 + 133 = 0.0073

RESULTS AND DISCUSSION

Since our Naïve bayes classifier has been trained using some already classified tweets, other incoming tweets not available in our collected dataset and not classified can now be trained using the Naïve Bayes Machine Learning Model. This can be done by passing the incoming tweet via the model and going through all the various stages mentioned earlier.

New Document

The Naïve Bayes Machine Learning Model has been trained with tweets and now ready to classify new tweets not present in our vocabulary. Depending on the words present in the incoming tweet, the existing dataset can be upgraded from time to time as any word not available can be added thereby increasing the size of the vocabulary. This is appropriate for an effective sentiment classification.

The battery has good quality©© is an example of a new Tweet that is not available in the collected dataset. This tweet is also coming in as unclassified and will be classified using the Naïve Bayes model.

Preprocess Document

The incoming tweet has to undergone cleaning to also remove irrelevant details like symbols, punctuation marks, and emojis which is important for accuracy in classification analysis. Words like * and © are removed since they lack proper English dictionary meaning. The incoming tweet becomes the battery has good quality after undergoing cleaning. Tokenization and vectorization also take place at this stage So that the model can recognize and effectively give prediction.

Train using the Naïve Bayes

The value of the naïve bayes classifier is the value that gives the highest number on computing the probability of both the positive and negative documents in the new dataset.

NB = MAX Val of P(Document+) x $P(W_j/+)$ and P(Document-) x $P(W_j/-)$

Where P (Document +) = the Probability of all the words in the positive document

And $P(W_j/+)$ = the Probability of each word present in the positive document depending on j which is the number of times the word occurred.

The battery has good quality

$P\ (Document/+)\ x\ P(wj/+)$

$P\ (Document/+) = \frac{10}{20} = 0.5$

$P(the+) = \frac{4+1}{57+133} = 0.0263$

$P(battery+) = \frac{1+1}{57+133} = 0.0105$

$P(has+) = \frac{1+1}{57+133} = 0.0105$

$P(good+) = \frac{3+1}{57+133} = 0.0211$

$P(quality+) = \frac{2+1}{57+133} = 0.0158$

0.5 x 0.0263 x 0.0105 x 0.0105 x 0.0211 x 0.0158 = 4.833 E^{-10} which is the result of probability for the positive document

$P\ (Document\ -/)\ x\ P(wj/-/)$

Where P (Document -) = the Probability of all the words in the negative document

And $P(W_j/-)$ = the Probability of each word present in the negative document depending on j which is the number of times the word occurred.

The battery has good quality

$P\ (Document/-) = \frac{10}{20} = 0.5$

$P(the-) = \frac{7+1}{141+133} = 0.0292$

$P(battery-) = \frac{0+1}{141+133} = 0.0036$

$P(has-) = \frac{0+1}{141+133} = 0.0036$

$P(good-) = \frac{0+1}{141+133} = 0.0036$

$P(quality-) = \frac{0+1}{141+133} = 0.0036$

Since the word battery, has, good and quality are not present in the negative vocabulary nk will be zero.

Therefore

$P\ (Document/-)\ x\ P(wj/-) =$

0.5 x 0.0292 x 0.0036 x 0.0036 x 0.0036 x 0.0036 = 2.452E^{-12} which is the result of probability for the negative document

Polarity Classification

Naive Bayes is very effective for text classification and is computationally efficient. Fron the incoming tweet: The battery has good quality will be classified as a positive tweet and given a score 1 as the result after the Naïve Bayes classification analysis for the P(Document+) is greater than the P(Document-). Any new word not available in the vocabulary can be added and updated as there is no restriction in the size of the dataset. This will thereby increase the number of unique words in the vocabulary for accurate classification.

CONCLUSION

Naïve bayes is a very effective machine learning tool for twitter product classification. It works using the mathematical independent assumption theory. Business and sole proprietor needs a means to detect the acceptability and growth of their products in the economy and this can be ascertained by analyzing the overall sentiments of various individuals using their products. In this work, the classification and analysis of twitter product using naïve bayes machine learning was conducted for mobile phone reviews. A total of twenty reviews were collected as sample comprising of ten positive tweets and ten negative tweets, neutral tweets were excluded because it's difficult to analyze the margin for effective sentiment analysis. Incoming reviews not present in the vocabulary were further trained using the model. The naïve bayes proves to be very effective for this analysis.

REFERENCES

Abinaya, R., Aishwaryaa, P., Baavana, S., and Selvi, N. T. (2016, July). Automatic sentiment analysis of user reviews. In *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, 158-162. IEEE.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. J. (2011). Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, 30-38.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., and Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. International Journal of Information Engineering and Electronic Business (IJIEEB), 8(4), 54-62, https://doi.org/10.5815/ijieeb.2016.04.07

Fang, X., and Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big data*, *2*, 1-14.

Hemalatha, I., Varma, G. P. S., and Govardhan, A. (2014). Automated sentiment analysis system using machine learning algorithms. *IJRCCT*, *3*(3), 300-303.

Korde, V., and Mahender, C. N. (2012). Text classification and classifiers: A survey. *International Journal of Artificial Intelligence and Applications*, *3*(2), 85.

Liu, Bing. (2012). Sentiment analysis and opinion mining, Morgan and Claypool publishers, May 2012.

Mohan, V., and Venu, S. H. (2016). Sentiment analysis applied to airline feedback to boost customers Endearment. *International Journal of Applied and Physical Sciences*, *2*(2), 51-58

Narayanan, V., Arora, I., and Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. In *Intelligent Data Engineering and Automated Learning–IDEAL 2013: 14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013. Proceedings 14*, 194-201, Springer Berlin Heidelberg.

Orie C.E, Egwali A.O. Amadin F. I. (2023). Machine Learning Based Sentiment of Tweet Data. 5th Annual Conference of the Society for Forensic and Analytical Scientist Nigeria.

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, *2*(1–2), 1-135.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques, Proc. 2002 Conf. on Empirical Methods in Natural Language Processing (EMNLP).

Prof. (Mr.) Prashant Kanade, Lakhan Rangwani, Pritee Wadhwa, Jitesh Watwani and Nitin Hazarani (2019). Automatic Sentiment Analysis of User Reviews, International Journal of Information and Computation Technology, 9(1), 5-10

Rai, A., and Jaiswal, K. (2017). Sentiment analysis by using unsupervised comment summarization. *International Research Journal of Engineering and Technology*, *4(5)*, 1105-1109.

Ramdhani, S. L., Andreswari, R., and Hasibuan, M. A. (2018, November). Sentiment analysis of product reviews using naive bayes algorithm: A case study. In *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 123-127, IEEE.

Shah, A. (2021). Sentiment analysis of product reviews using supervised learning. *Reliability: Theory and Applications*, *16*(1), 243-253.

Shukla, R., and Mishra, N. (2016). Framework for Sentiment Analysis of Twitter Post. International Journal of Innovative Research in Science, Engineering and Technology, 5(3).

Singla, Z., Randhawa, S., and Jain, S. (2017). Sentiment analysis of customer product reviews using machine learning. In *2017 international conference on intelligent computing and control (I2C2)*, 1-5, IEEE.

Suchdev, R., Kotkar, P., Ravindran, R., and Swamy, S. (2014). Twitter sentiment analysis using machine learning and knowledge-based approach. *International Journal of Computer Applications*, *103*(4).

Umarani, V., Julian, A. and Deepa, J. (2021). Sentiment Analysis using various Machine Learning and Deep Learning Techniques. Journal of the Nigerian Society of Physical Sciences. Volume 3, issue 4, pp 385-394.

Vidhya, K. A., and Aghila, G. (2010, February). Hybrid text mining model for document classification. In *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, 1, 210-214, IEEE.