Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 5, July 2021 : 2817 - 2831

Research Article

# An Approach for Sentiment Analysis of Customer Review Data

**[1]Deevi Radha Rani, [2]Ongole Gandhi, [3]Viswandapalli Anusha, [4]Shaik Shabbir Hussain, [5]Divya Vadlamudi**

[1,2,3,4]Department of CSE, VFSTR Deemed to be University, Vadlamudi, Guntur, AP, India
[5]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
[1]dradharani29@gmail.com, [2]ongolegandhi@gmail.com, [3]anusha6785@gmail.com, [4]sksh_cse@vignan.ac.in, [5]divya.movva@kluniversity.in

## Abstract

Sentiment analysis is more promising tool to select the best product based on the customer reviews. The increase in number of websites and brands who are advertising their products leads to increase of customer reviews day by day. Manually it is not feasible to analyze and decide the opinion against a product using those huge reviews. Sentiment analysis automates the process of classifying the products as positive, negative and neutral based on customer reviews. This paper focus on performing sentiment analysis on the text data that contains the customer reviews to obtain the sentiment i.e., opinion of the user about the product from the reviews that the customer have given. This paper also presents the classification of sentiment analysis techniques and stages in sentiment analysis. The approach used in this paper uses both lexicon based technique and machine learning technique, especially SVM. Performance of our proposed approach is evaluated using precision, recall and F1-score. The accuracy of different decisions is also calculated. We used kaggle dataset for the experimenting our proposed sentiment analysis approach.

**Keywords**: Sentiment Analysis, Customer Reviews, Preprocessing, Support Vector Machine, NLTK, TextBlob, Kaggle Dataset;

## Introduction

Sentiment is an attitude driven by feeling or judgment. Sentiment analysis can also be termed as opinion mining. The customers are free to post their own opinions on various e-commerce sites but quality of those opinions is not guaranteed. So sentiment analysis is more challenging. Sentiment analysis is the process of gathering opinions of an individual customer or a brand's audience in communication with a customer service representative. Sentiment analysis helps determine conversations, language, and voice inflections to calculate emotions related to a business, product, or brand. Sentiment analysis has its hands in numerous areas. For instance,

sentiment analysis can be performed on social media to determine sentiments of customers on a trending topic.

Besides, sentiment analysis can be used by call centers to monitor customer support performance. Furthermore, it has its hands in politics to gather reviews about policy changes, campaign announcements, and many others. Many of the products and services that we are using these days are not meeting the customer expectations resulting in the customer dissatisfaction and decreased sales in the products. And many other customers want to know the working of the product before they purchase to know whether the product worth the money they spent. These cases are provided with a solution by the Sentiment Analysis by calculating the Sentiment of user opinions about the product. This paper focus on sentiment analysis on Amazon cell phone reviews dataset in order to determine the opinion (sentiment i.e. positive, negative or neutral) of users of cell phones so that business people and other customers can get an overview regarding a particular product. In this paper we used TextBlob and NLTK for calculating sentiment score and SVM for classification of customer reviews.

The rest of the paper is organized as follows, related work section discusses the importance, types and classification of sentiment analysis; proposed approach section gives the stages of sentiment analysis; experimental results and discussion section presents the dataset used and results obtained from our implementation; finally the paper concludes in conclusion section.

## Related Work

Sentiment analysis has been an important tool for brands looking to learn more about how their customers are thinking and feeling. It is a relatively simplistic form of analytics that helps brands find key areas of weakness (negative sentiments) and strengths (positive sentiments). Moving forward, sentiment analysis is finding a place in other organizations. During Brexit and the 2016 US election, these data tools were used to measure emotions and attempt to predict the outcome of these events. This has led to non-brand organizations turning to sentiment analysis for their own needs. Additionally, the insights gained from these tools are becoming much deeper, as a result of emerging social media platforms and features. Sentiment analysis simply looks more popular in the future.

A. Why Sentiment Analysis?

Sentiment analysis on customer review data is an important tool to get positive outcomes with customer interactions [2]. Sentiment analysis can be done manually or using any technique described in this paper, the sentiment analysis works as described below:

a) Proactive business operation: Sentiment analysis provides useful insights for the analysis of events and adaptable categories. In order to analyze customer attitudes towards a particular topic or product, marketers can obtain information from blogs, reviews, social media postings. In addition, marketers can recognize purchasing intent

through the sales funnel, which lets them determine the new brand sentiment-influencing promotions and also identify the segment that draws greater interest.

b) Deep Audience insight: Sentiment Analysis helps to provide helpful and better customer insights that drive marketers to curate future content and campaign plans. It also helps marketers address market research by tailoring their marketing tone before further moving with new product features.
c) Better ROI on marketing campaign: Sentiment analysis helps to assess the amount of positive and negative conversations about a particular brand or product that helps marketers succeed in their marketing campaign. In addition, it helps in the marketing campaign to scale the ROI by combining numerical and non-numeric data.
d) Improved customer service: Sentiment analysis is seen as an effective tool for understanding the behavior of a customer and monitoring their dissatisfaction, if any. It also helps a poor customer rating to be changed to nice. This helps marketers push their business to an entirely different success level.
e) Good public relation practice: Analysis of feedback encourages exposure to understanding of the brand, which is a public relations must. It also helps marketers analyze how relevant and informative each message is to their customers. Analysis of feelings helps to identify how the public feels about certain topics.

B. Types of Sentiment Analysis

i) Document Level: Typically this type of sentiment analysis is applied over the entity which helps to recognize the negative or positive views of an individual entity by using documents [1] [6].
ii) Comparative Level: The main purpose of this type of sentiment analysis is to identify the opinions using comparative sentence [1] [6].
iii) Aspect based: This type of sentiment analysis analyzes multiple entities. In the type of sentiment analysis, it typically helps to evaluate the negative and positive item-based aspect [1] [6].

C. Classification of Sentiment Analysis Techniques

Sentiment Analysis techniques can be generally classified as lexicon based techniques, machine learning techniques and hybrid techniques. Lexicon based techniques finds the lexicon to analyze the customer reviews. Dictionary based approach finds the words in the customer reviews and then search for the meaning where as corpus based approach works statistically and semantically. Machine learning techniques are broadly classified as unsupervised learning and supervised learning. In supervised learning various classifiers are presented by various authors to classify the customer reviews like linear classifier, rule based classifier, decision tree classifier, probabilistic classifier. Linear classifiers like SVM [7], Neural Networks [8] and probabilistic classifiers like navie bayes [9], Bayesian network [10], maximum entropy [11] can be used for sentiment classification.

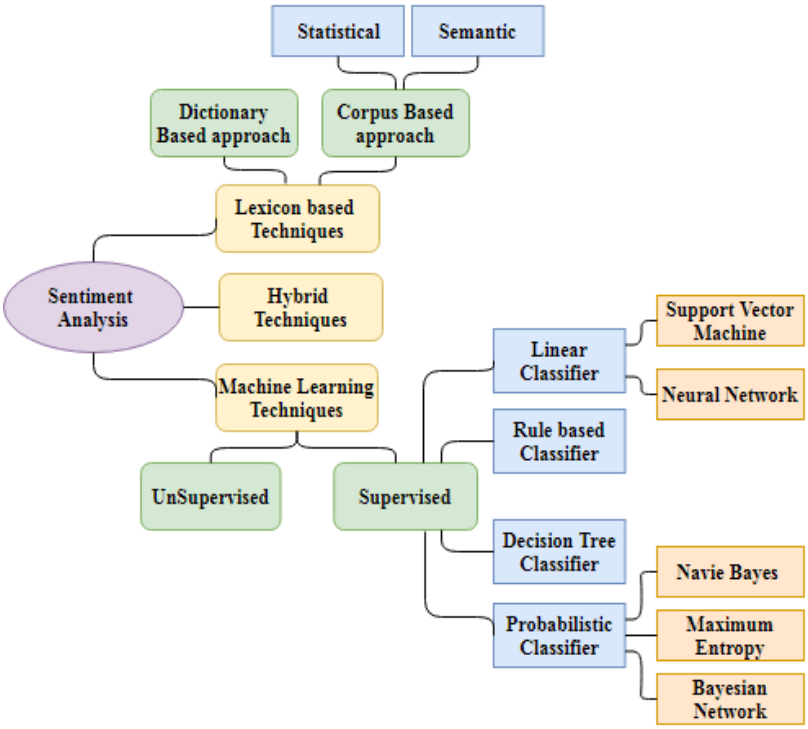


Figure 1 Classification of Sentiment Analysis Techniques

## The Proposed Approach of Sentiment Analysis

As shown in Figure 2, the proposed method for sentiment analysis consists of following stages: Data collection, Preprocessing, Feature Selection, Sentiment Classification. Each step of the proposed method is explained below:

### 1. Data Collection

The data on which we are going to perform the sentiment Analysis is the textual data that contains the reviews given to the product, details of the reviewer and the product. We can create our own datasets with the requirements we have or we can also use the existing datasets that are available on internet, one such resource for datasets is Kaggle Datasets.

### 2. Preprocessing the Data

The data we collect from the users in the form of reviews contains a lot of noise that reduces the accuracy of the sentiment that we have calculated for the reviews. Noise is anything that is not useful for our process of calculating sentiment one such example of noise is emoji's that are given by users in the reviews. So we have to remove these kinds of noises from the data for accurate sentiment results. This is called as pre-processing or cleaning the data. Preprocessing involves 3 steps:

Step 1: Removal of Punctuation:

For same set of words over punctuation may change the sentiment of the sentence. So, punctuation need to be eliminated. Punctuation that is to be removed is "! " # $ % & ' ( ) * + , - . / : ; < = > ? @ [ /] ^ _` {|} ~"

Step 2: Removal of other special characters like emoji's:

While giving reviews users may also use emoji symbols with which we can't generate a semantic score using Sentiment Analyzer. So those symbols need to be removed. These symbols are removed by encoding text first into ascii standard and again decoding into ascii standards. Because ascii standard contains only 128 unique characters in which emoji's are not available. So, during encoding these symbols were simply ignored.

Step 3: Removal of Stop words:

A stop word is a commonly used word such as "the"," a"," an"," is" etc. that need to be ignored during pre-processing of data because of their frequency of occurrence in text they consume much memory and processing resources. So, to reduce this overhead problem stop words are need to be eliminated.

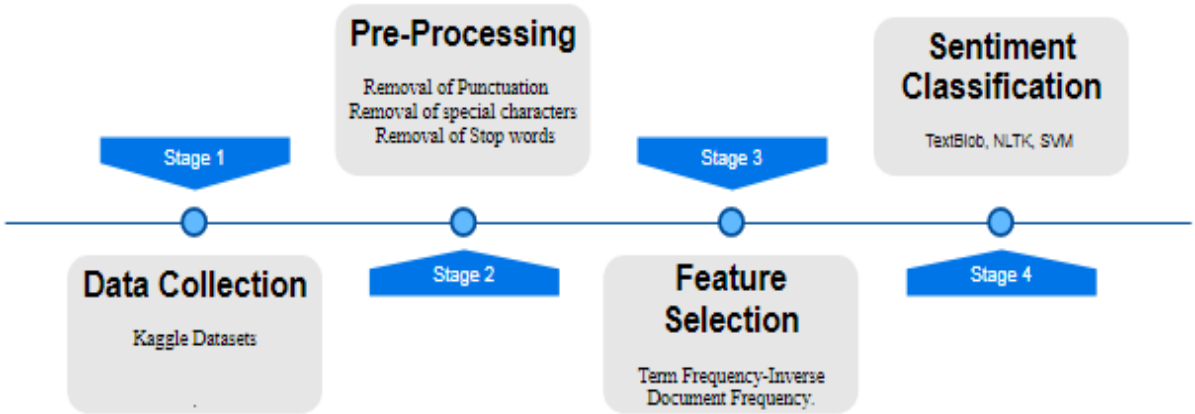


Figure 2 Stages in Sentiment Analysis

## 3. Feature Selection

Term Frequency-Inverse Document Frequency (TF-IDF) is most used text mining approach [3]. TF-IDF is a weight metric which decide the importance of word in a review. Term Frequency (TF) measures the occurrence of term t in a review r. TF is calculated as

$$\mathrm{TF}(t,r) = \frac{\text{No. of occurences of term } t \text{ in review } r}{\text{No. of terms in review } r}$$

Inverse Document Frequency (IDF) measures the importance of a term. IDF is calculated as

$$\mathrm{IDF}(t) = \log_e \frac{\text{Total no. of reviews}}{\text{Total no. of reviews with term } t}$$

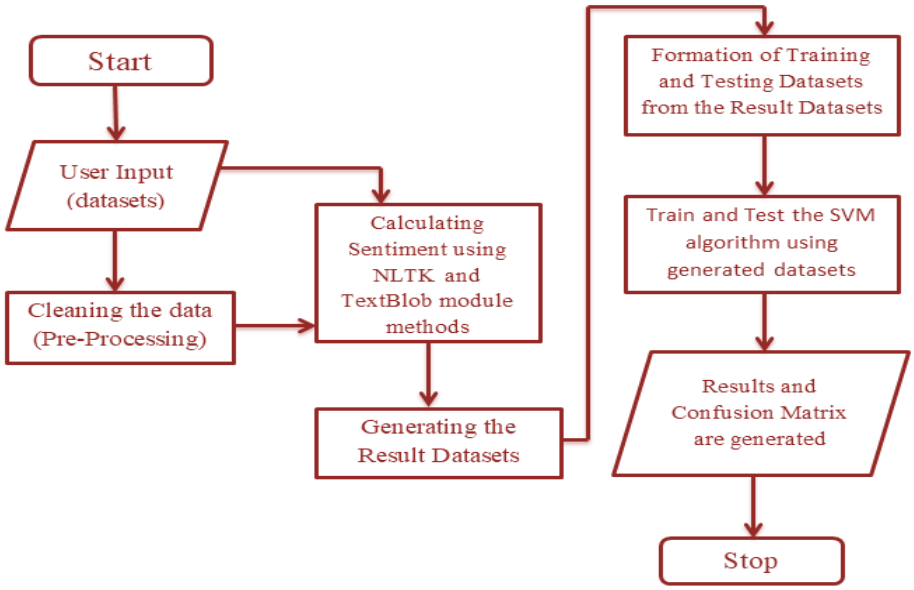


Figure 3 Flow diagram of the proposed approach for sentiment analysis

## 4. Sentiment Calculation

The useful words of the text data that we have collected are considered as tokens and these tokens enable us to calculate sentiment of the review. In this paper we used NLTK and TestBlob to calculate the sentiment score. Training dataset and testing dataset is generated using NLTK and TestBlob for classification of sentiment. Support Vector Machine (SVM) [5] is also used for sentiment classification and proposed approach is evaluated based on various parameters. Figure 3 described the flow of proposed approach for sentiment analysis.

## Experimental Results and Discussion

### A. Dataset

The dataset used for the experimental purpose is Kaggle Dataset. It is simplest and best-supported file type available on Kaggle is the “Comma-Separated List”, or CSV for tabular data. CSVs uploaded to Kaggle should have a header row consisting of human-readable field names. A CSV representation of a shopping list with a header row, for example, looks like this:

*id, type, quantity*
*0, bananas, 12*

CSV files will also have associated column descriptions and column metadata. The column descriptions allows you to assign descriptions to individual columns of the dataset, making it easier for users to understand what each column means. Column metrics, meanwhile, present high-level metrics about individual columns in a graphic format. Table 1 dataset contains the information regarding the products i.e. Cellphones. Table 2 dataset contains information regarding the reviewer and review.

Table 1 Items.csv Dataset

| Column Name | Column Description |
|---|---|
| asin | Amazon Standard Identification Number (10-digited) used for product-identification within Amazon.com |
| brand | Name of the product brand |
| title | Product title |
| url | Product URL |
| image | Product image URL |
| rating | Product average rating value |
| reviewUrl | Product Review page URL |
| totalReviews | Total number of product reviews |
| Price(in US$) | Price of product in US $ |

Metadata regarding Items.csv: Total Number of columns: 9, Total Number of rows: 792

Table 2 Reviews.csv Dataset

| Column Name | Column Description |
|---|---|
| asin | Amazon Standard Identification Number (10-digited) used for product-identification within Amazon.com |
| name | Reviewer name |
| rating | Reviewer rating (scale 1 to 5) |
| date | Review date |
| verified | Valid customer (TRUE or FALSE) |
| title | Review title |
| body | Review content |
| helpfulVotes | Helpful feedbacks |

Metadata regarding reviews.csv: Total Number of columns: 8, Total Number of rows: 82815

## B. Implementation

From the two datasets we don't require all attributes because at this stage we're interested in understanding the basic procedure do sentiment analysis using pre-defined Python NLTK (Natural Language processing Toolkit) module.

In future based on requirement we may include other attributes that have impact in predicting sentiment. At this stage 1 we're going to get predict sentiment at each single review. We are not predicting the sentiment of the product based on all its related reviews and attributes.

So, for now we are choosing asin and body attributes from the review.csv dataset for our analysis. The body attribute is selected because it specifies the content of the review and asin is selected because it is common attribute in both datasets. But for now, it is of no use but in future we may require it to map reviews to particular product. No attribute is selected from items.csv data because it is completely about product information but right now, we interested in reviews so items.csv is not considered.

Now we are going to perform sentiment analysis in two different cases i.e. Sentiment analysis using cleaned dataset and Sentiment Analysis using uncleaned dataset. For both the case calculating polarity score (sentiment score) is same but difference is that in case of cleaned dataset, the dataset undergoes pre-processing i.e. cleaning of data whereas in uncleaned dataset, the same dataset is passed as input to Sentiment Analyzer to calculate sentiment or polarity scores for each review. Removing stop words, nltk.stopwords("english") provides a corpus that contain "not" as one of the stop word. To remove punctuation using Python we use punctuation in string module. Figure 4, Figure 5, Figure 6 presents the screenshot for 3 steps of preprocessing respectively.

```
In [5]: runfile('C:/Users/19PJ283/.spyder-py3/tring.py', wdir='C:/Users/19PJ283/.spyder-
py3')

With Punctuation:The phone "was great" until the pixels started decaying. Now the whole
screen is purple!! Cant used the cell anymore, that's the risk you run with used phones.
Without Punctuation: the phone was great until the pixels started decaying now the whole
screen is purple cant used the cell anymore thats the risk you run with used phones
```

Figure 4 Removal of punctuation

```
In [6]: runfile('C:/Users/19PJ283/.spyder-py3/tring.py', wdir='C:/Users/19PJ283/.spyder-
py3')

With symbols:I wanted black but received blue. And I wanted it for Verizon but it isnâ€™t
compatible ðŸ¤¦ðŸ»â€â™€ï¸
Without symbols: i wanted black but received blue. and i wanted it for verizon but it isnt
compatible
```

Figure 5 Removal of emoji/symbol

```
In [10]: runfile('C:/Users/19PJ283/.spyder-py3/tring.py', wdir='C:/Users/19PJ283/.spyder-
py3')

with stopwords:I was disappointed becouse I buy a new phone and they send me second hand
phone but it works ok for now
without stopwords: disappointed becouse buy new phone send second hand phone works ok
```

Figure 6 Removal of stop words

For calculating sentiment score using Python we use a module called nltk. To calculate polarity score an object for class SentimentIntensityAnalyzer to be initialized which has to be imported from nltk.sentiment.vader. Then using polarity_score() method which is available in SentimentIntensityAnalyzer class and passing text as argument to this method scores are calculated. The output of polarity_score() method is a dictionary data type containing negative, positive, neutral and compound polarity scores. Compound polarity score measure the overall sentiment of text. Based on compound value text is determined positive (compound value>0), negative (compound value<0) or neutral (compound value = 0).

For calculating sentiment score using Python we use a module called textblob. To calculate polarity score an object for class TextBlob to be initialized which has to be imported from textblob module and target text has to be passed as parameter to TextBlob. Then using sentiment() method which is available in TextBlob class polarity score is calculated. The output of sentiment() method is a tuple data type containing Polarity score and subjectiviy score. Based on compound value text is determined positive (polarity score>0), negative (polarity score <0) or neutral (polarity score = 0).

During polarity scores calculation all the results are stored in list data type. In cleaned dataset case, cleaned reviews are also provided in dataset under attribute cleaned review and labels are specified under the attribute decision along with asin and review attributes. In uncleaned dataset case, labels are provided under the attribute decision along with other attributes in the figure 7.

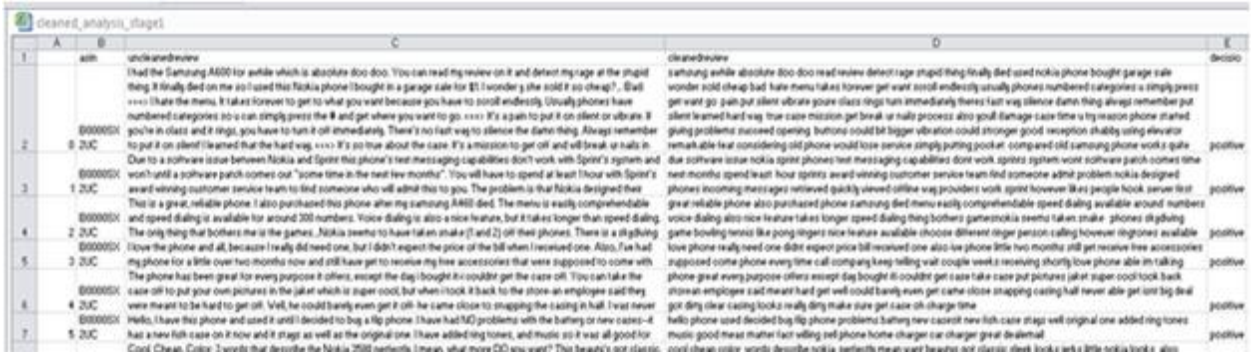

Figure 7 Cleaned test case: cleaned_analysis_stage1.csv

| | A | B | C | D |
|---|---|---|---|---|
| 1 | | asin | review | decision |
| 2 | 0 | B00005X2UC | I had the Samsung A600 for awhile which is absolute doo doo. You can read my review on it and detect my rage at the stupid thing. It finally died on me so I used this Nokia phone I bought in a garage sale for $1. I wonder y she sold it so cheap?... Bad: ==> I hate the menu. It takes forever to get to what you want because you have to scroll endlessly. Usually phones have numbered categories so u can simply press the # and get where you want to go. ==> It's a pain to put it on silent or vibrate. If you're in class and it rings, you have to turn it off immediately. There's no fast way to silence the damn thing. Always remember to put it on silent! I learned that the hard way. ==> It's so true about the case. It's a mission to get off and will break ur nails in the process. Also, you'll damage the case each time u try. For some reason the phone started giving me problems once I did succeed in opening it. ==> Buttons could be a bit bigger. Vibration could be stronger. Good: ==> Reception is not too shabby. I was using it in the elevator which is a remarkable feat considering my old phone would lose service by simply putting it in my | positive |
| 3 | 1 | B00005X2UC | Due to a software issue between Nokia and Sprint this phone's text messaging capabilities don't work with Sprint's system and won't until a software patch comes out "some time in the next few months". You will have to spend at least 1 hour with Sprint's award winning customer service team to find someone who will admit this to you. The problem is that Nokia designed their phones so that incoming messages are retrieved quickly and then viewed "offline" the way most providers work. Sprint, however, likes to have people hook up to their server first and then | positive |
| 4 | 2 | B00005X2UC | This is a great, reliable phone. I also purchased this phone after my samsung A460 died. The menu is easily comprehendable and speed dialing is available for around 300 numbers. Voice dialing is also a nice feature, but it takes longer than speed dialing. The only thing that bothers me is the games...Nokia seems to have taken snake (1 and 2) off their phones. There is a skydiving game, bowling, and tennis (like pong). The ringers are very nice, and a feature is available to choose a different ringer for each person calling. However, ringtones are not available | positive |

Figure 8 Uncleaned test case: uncleaned_analysis_stage1.csv

In stage 2 we considered each review and its sentiment predicted and stored in the corresponding files. From the input dataset reviews.csv we observed that for a unique product having multiple reviews. This means for a single unique product multiple review are there. So here using asin attribute as key we are going to find total number of reviews on that product, again among them total number of positive reviews, total number of negative reviews, total number of neutral reviews and total number of no reviews. Again, in stage 2 we have two cases cleaned and uncleaned. In both the case whole process is same but input datasets are different.

For cleaned case, Input dataset is cleaned_analysis_stage1.csv and Output dataset is cleaned_analysis_stage2.csv as shown in figure 9.

cleaned_analysis_stage2

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | asin | brand | product_title | total reviews | tot_positive | tot_negative | tot_neutral | tot_noreview |
| 2 | 0 | B0000SX2 | Nokia | Dual-Band / Tri-Mode Sprint PCS Phone w/ Voice Activated Dialing & Bright Whit | 14 | 13 | 1 | 0 | 0 |
| 3 | 1 | B0009N5L | Motorola | Motorola i265 phone | 7 | 4 | 2 | 1 | 0 |
| 4 | 2 | B0005KTZ | Motorola | MOTOROLA C168i AT&T CINGULAR PREPAID GOPHONE CELL PHONE | 22 | 16 | 3 | 3 | 0 |
| 5 | 3 | B00198M | Nokia | Nokia 6500 Slide Black/silver Unlocked Cell Phone | 5 | 2 | 3 | 0 | 0 |
| 6 | 4 | B001AO4( | Motorola | Motorola i335 Cell Phone Boost Mobile | 21 | 12 | 9 | 0 | 0 |
| 7 | 5 | B001DCJA | Motorola | Motorola V365 no contract cellular phone AT&T | 12 | 10 | 1 | 1 | 0 |
| 8 | 6 | B001DZY4 | Sony | Sony Ericsson G700 Triband GSM Phone Bronze (Unlocked) | 1 | 1 | 0 | 0 | 0 |
| 9 | 7 | B001GQ3I | Nokia | Nokia 1680 Black Phone (T-Mobile) | 3 | 2 | 1 | 0 | 0 |
| 10 | 8 | B0027VKC | Nokia | Nokia New 1100 for Tracfone | 8 | 7 | 1 | 0 | 0 |
| 11 | 9 | B00280QJ | Samsung | Samsung T301G Prepaid Phone (Tracfone) | 133 | 98 | 27 | 8 | 0 |
| 12 | 10 | B0029X7L | Motorola | Motorola i205 cell phone nextel/Boost | 2 | 0 | 1 | 1 | 0 |

Figure 9 Output dataset: cleaned_analysis_stage2.csv

For uncleaned case, Input dataset is uncleaned_analysis_stage1.csv and Output dataset is uncleaned_analysis_stage2.csv as shown in figure 10.

uncleaned_analysis_stage2

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | asin | brand | product_title | total reviews | tot_positive | tot_negative | tot_neutral | tot_noreview |
| 2 | 0 | B0000SX2UC | Nokia | Dual-Band / Tri-Mode Sprint PCS Phone w/ Voice Activated Dialing & Bright White Backlit | 14 | 13 | 1 | 0 | 0 |
| 3 | 1 | B0009N5L7K | Motorola | Motorola i265 phone | 7 | 3 | 3 | 1 | 0 |
| 4 | 2 | B0005KTZD5 | Motorola | MOTOROLA C168i AT&T CINGULAR PREPAID GOPHONE CELL PHONE | 22 | 15 | 3 | 4 | 0 |
| 5 | 3 | B00198M12M | Nokia | Nokia 6500 Slide Black/silver Unlocked Cell Phone | 5 | 2 | 3 | 0 | 0 |
| 6 | 4 | B001AO4OUC | Motorola | Motorola i335 Cell Phone Boost Mobile | 21 | 11 | 10 | 0 | 0 |
| 7 | 5 | B001DCJAIG | Motorola | Motorola V365 no contract cellular phone AT&T | 12 | 9 | 2 | 1 | 0 |
| 8 | 6 | B001DZY4KI | Sony | Sony Ericsson G700 Triband GSM Phone Bronze (Unlocked) | 1 | 0 | 1 | 0 | 0 |
| 9 | 7 | B001GQ3DIM | Nokia | Nokia 1680 Black Phone (T-Mobile) | 3 | 3 | 0 | 0 | 0 |
| 10 | 8 | B0027VKQPE | Nokia | Nokia New 1100 for Tracfone | 8 | 7 | 1 | 0 | 0 |
| 11 | 9 | B00280QJFU | Samsung | Samsung T301G Prepaid Phone (Tracfone) | 133 | 89 | 39 | 5 | 0 |
| 12 | 10 | B0029X7UHC | Motorola | Motorola i205 cell phone nextel/Boost | 2 | 0 | 1 | 1 | 0 |

Figure 10 Output dataset: uncleaned_analysis_stage2.csv

Table 3 describes the metadata on output datasets.

Table 3 Metadata of output datasets

| Column name | Column Description |
|---|---|
| S.No | Specifies serial number |
| asin | Amazon Standard Identification Number(key attribute) |
| brand | Specifies mobile brand name |
| Product_title | Specifies short description regarding mobile |
| total reviews | Specifies total reviews available on that particular product |
| tot_positive | Specifies total positive reviews on that particular product |
| tot_negative | Specifies total negative reviews on that particular product |
| tot_neutral | Specifies total neutral reviews on that particular product |
| tot_noreviews | Specifies total no reviews on that particular product |

During polarity scores calculation all the results are stored in list data type. In cleaned dataset case, cleaned reviews are also provided in dataset under attribute cleaned review and labels are specified under the attribute decision along with asin and review attributes. In uncleaned dataset case, labels are provided under the attribute decision along with other attributes. Table 4 and Table 5 gives the sentiment analysis using NLTK and TextBlob respectively. Figure 11 and Figure 12 shows the sentiment analysis using NLTK and TextBlob respectively.

Table 4 Sentiment Analysis for NLTK

| Labels | UnCleaned | Cleaned |
|---|---|---|
| Positive | 55962 | 59667 |
| Negative | 17157 | 11449 |
| Neutral | 9676 | 11507 |
| Empty Review Empty Judgement | 20 | 192 |
| TOTAL | 82815 | 82815 |

Table 5 Sentiment Analysis for TextBlob

| Labels | UnCleaned | Cleaned |
|---|---|---|
| Positive | 60040 | 62177 |
| Negative | 10801 | 9904 |
| Neutral | 11954 | 10542 |
| Empty Review Empty Judgement | 20 | 192 |
| TOTAL | 82815 | 82815 |

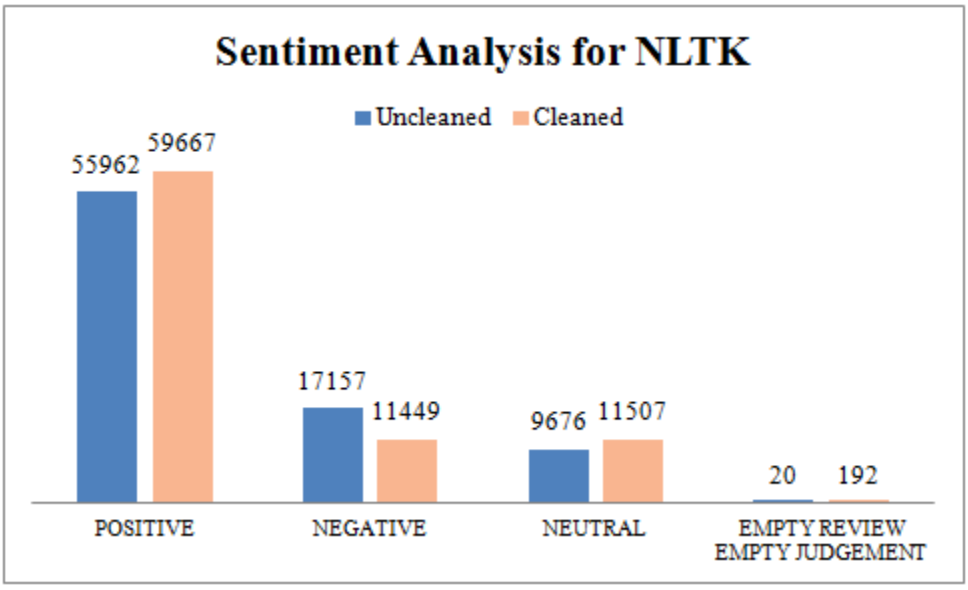


Figure 11 Sentiment Analysis for NLTK

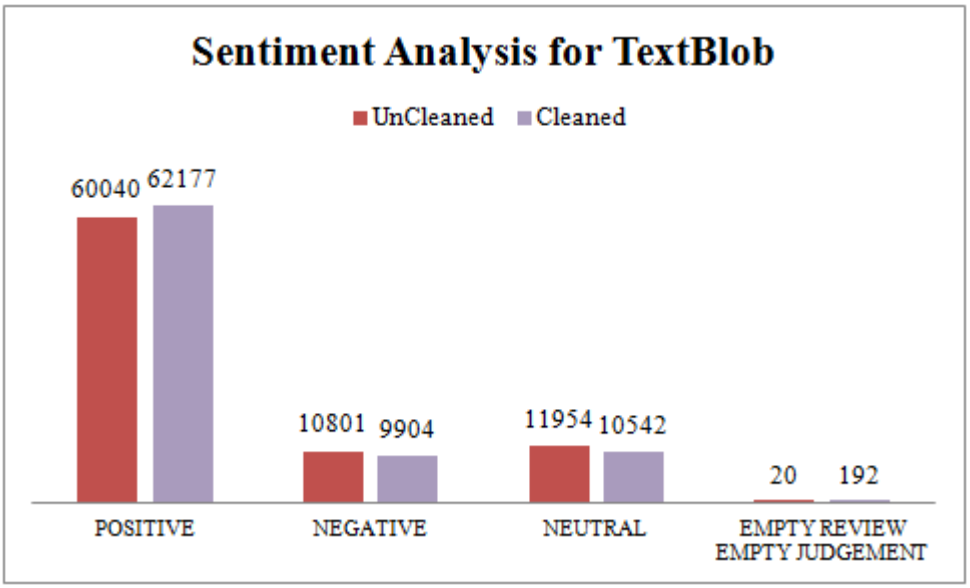


Figure 12 Sentiment Analysis for TextBlob

Support Vector Machine is used for classification of sentiment after calculating the sentiment score using NLTK and TextBlob. Classification_report() method that is used in our program returns a dictionary dataset with elements precision, recall, f-score and support. These are all calculated for positive, negative, neutral and empty review and empty judgment categories. SVM has defined input and output format. Input is a vector space and output is zero or one (positive/negative). Performance of our proposed approach is evaluated using precision, recall and F1-score. Precision calculates how many selected items are relevant, it is the ratio of number of items correctly labeled as positive to total number of positively classified items. Recall calculates how many relevant items are selected, it is the ratio of total number of positively labeled items to total items which are truly positive. F1-score is harmonic mean of precision and recall. Figure 13 and Figure 14 represents the performance evaluation of SVM for NLTK and TextBlob respectively.

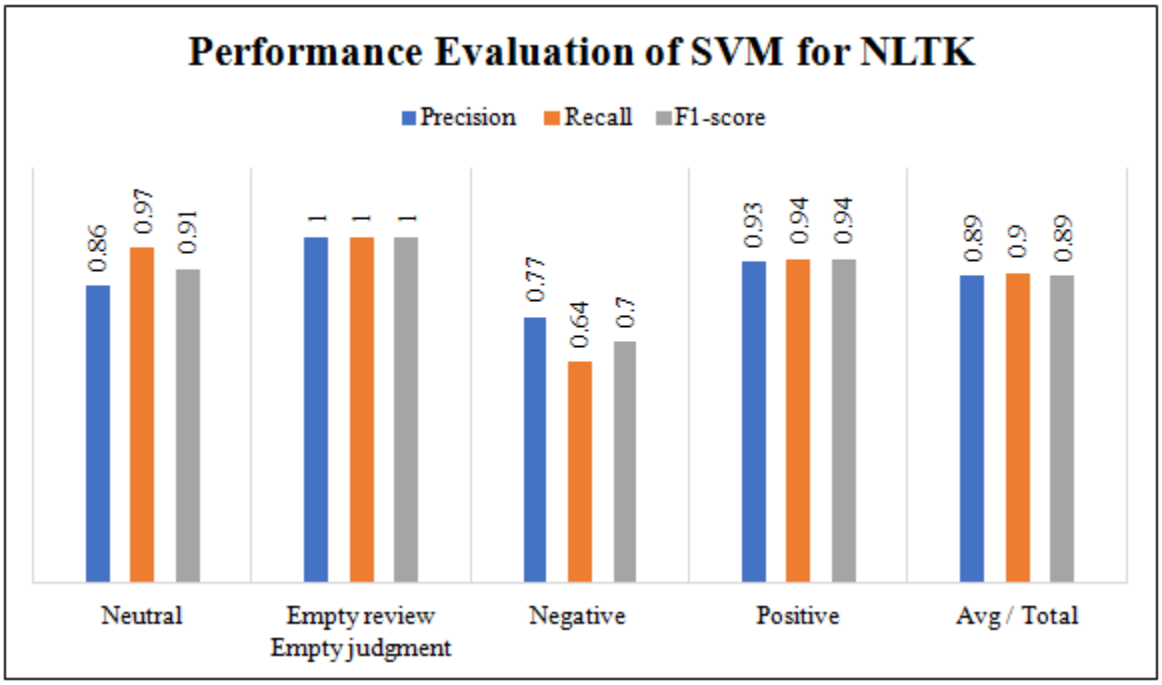


Figure 13 Performance evaluation of SVM for NLTK

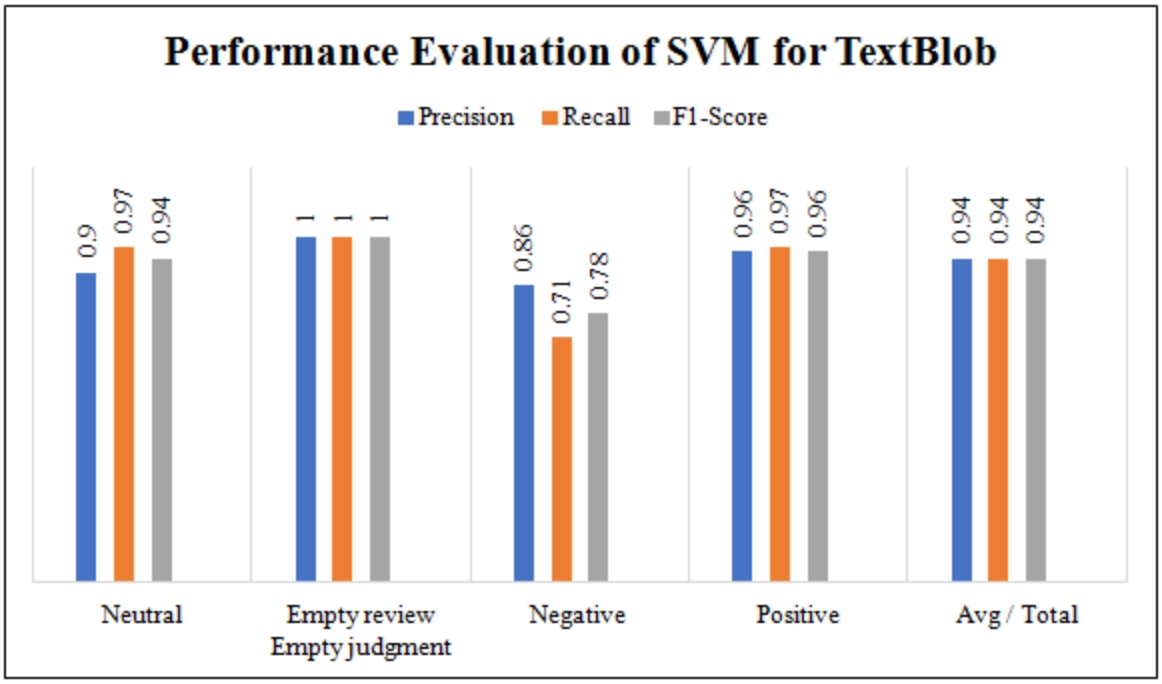


Figure 14 Performance evaluation of SVM for TextBlob

Accuracy is a performance measure for classification. It calculates how many items are predicted corrected. The accuracy analysis for NLTK and Texblob on various decisions are shown in the Figure 15.

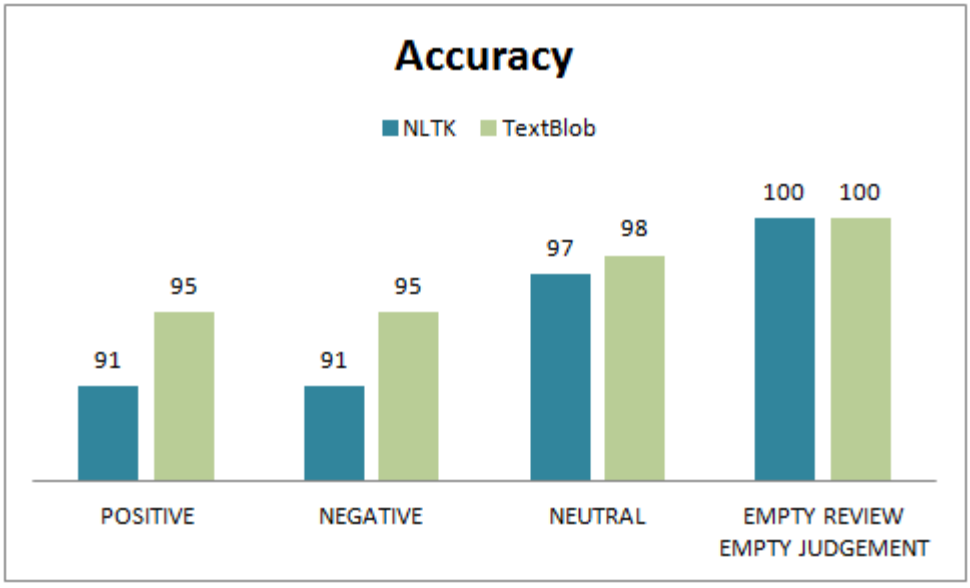


Figure 15 Accuracy analysis for NLTK and Texblob

## Conclusion

Many of the products and services that we are using these days are not meeting the customer expectations resulting in the customer dissatisfaction and decreased sales in the products. Sentiment analysis helps to select the best product we buy. Many researchers analyzed the product reviews in various ways that we described in this paper. Our proposed approach has given 86% precision with SVM for NLTK and 90% precision with SVM for TextBlob. The accuracy of different decisions is also calculated and gives 95% accuracy for positive and negative decision. As a future work we propose to analyze customer reviews further based on a specific brand and implement other techniques of sentiment analysis for calculating sentiment score and classification techniques.

## References

1. Menaria, Hemant Kumar, Pritesh Nagar, and Mayank Patel. "Tweet Sentiment Classification by Semantic and Frequency Base Features Using Hybrid Classifier." *First International Conference on Sustainable Technologies for Computational Intelligence*. Springer, Singapore, 2020.
2. Naveen Joshi, 5 reasons why you must care about customer sentiment analysis [online] 13, March 2018, https://www.allerin.com/blog/5-reasons-why-you-must-care-about-customer-sentiment-analysis, (Accessed 25, March, 2020).
3. Avinash, M., and E. Sivasankar. "A Study of Feature Extraction Techniques for Sentiment Analysis." *Emerging Technologies in Data Mining and Information Security*. Springer, Singapore, 2019. 475-486.

4. Avinash, M., and E. Sivasankar. "A Study of Feature Extraction Techniques for Sentiment Analysis." *Emerging Technologies in Data Mining and Information Security*. Springer, Singapore, 2019. 475-486.
5. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5.4 (2014): 1093-1113.
6. Collomb, Anaıs, et al. "A study and comparison of sentiment analysis methods for reputation evaluation." *Rapport de recherche RR-LIRIS-2014-002* (2014).
7. Luo, Fang, Cheng Li, and Zehui Cao. "Affective-feature-based sentiment analysis using SVM classifier." *2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2016.
8. Severyn, Aliaksei, and Alessandro Moschitti. "Twitter sentiment analysis with deep convolutional neural networks." *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2015.
9. Dey, Lopamudra, et al. "Sentiment analysis of review datasets using naive bayes and k-nn classifier." *arXiv preprint arXiv:1610.09982* (2016).
10. Chaturvedi, Iti, et al. "Bayesian network based extreme learning machine for subjectivity detection." *Journal of The Franklin Institute* 355.4 (2018): 1780-1797.
11. Lee, Huey Yee, and Hemnaath Renganathan. "Chinese sentiment analysis using maximum entropy." *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*. 2011.