BOOK CHAPTER | Man-In-The-Middle-Attacks

# Sentiment Analysis of Tweets Using Natural Language Processing (An Outline)

[1]Adepegba Oluwafunmilola, [1] Onwukwe David, [2]Amao Folake (P.hD) & [3]Adepegba Solomon
[1]Department of Computer Science, Adeleke University Ede, Nigeria
[2]Department of Mathematics, Adeleke University Ede, Nigeria
[3]Department of Computer Science, University of Ibadan, Ibadan, Nigeria
**E-mail:** [1]adepegbafunmilola@adelekeuniversity.edu.ng

## Abstract

The computational study of people's statements, opinions, sentiments, and emotions expressed in written language is known as sentiment analysis (SA), or opinion mining. It is the most well-known aspect of natural language processing, in which we try to figure out whether a piece of text contains subjective information, such as whether the attitude expressed in the text is positive, negative, or neutral. Here, an overview of what sentiment analysis is all about, is outlined by reviewing past related works on sentiment analysis and how it can be applied in our modern-day world making use of one of the most popular social media platforms known as Twitter which offers organizations a fast and effective way to analyze customers' perspectives towards improving on the products and services they render thereby maximizing profit or success in their respective fields. The study reports on building a sentiment analysis classifier making use of natural language processing in machine learning and other tools to build a web application that would enable a user to enter sample tweets and view the polarity. The web application would also be used in extracting a vast number of tweets at a time from Twitter and then the results classify users' perspectives on a particular subject or hashtag into positive and negative.

**Keywords:** Sentiment Analysis, Opinion Analysis, Twitter, Tweets, Natural language processing

## Introduction

A man-in-the-middle also known as monster-in-the-middle(Gabbi and Luke, 2019), monkey-in-the-middle, machine-in-the-middle (John, 2019), meddler-in-the-middle or person-in-the-middle (ACSC, 2020) attack is a cyberattack in which the perpetrator discreetly relays and perhaps modifies messages between two users who feel they are speaking directly with each other while the perpetrator has positioned himself between them (Elakrat et al., 2018). A laudable example of mam-in-the-middle-attack is eavesdropping, where the attacker establishes separate connections with the victims and send and receive data between them.



**Citation**: Adepegba O., Onwukwe, D. Amao, F. & Adepegba, S. (2022).
Sentiment Analysis of Tweets Using Natural Language Processing (An Outline)
SMART-IEEE-ACity-ICTU-CRACC-ICTU-Foundations Series

## Introduction

The worldwide use of social media has enabled the automatic acquisition and analysis of social media data for the finding of intriguing patterns and information within these data sets." (Adebisi and colleagues, 2018) People don't just post pictures and talk about their lives and events on social media sites and apps like Facebook, Twitter, Instagram, and Snapchat; they also share opinions about the products they've just purchased and the vendor. In addition they share information on the services they've just paid for, their political views and opinions about political leaders, and so on. And these opinions cannot be readily dismissed, as they have a significant impact on the image of the people concerned

## Natural Language Processing

Natural Language Processing (NLP) is a branch of Artificial Intelligence that allows computers to understand, perceive and manipulate human language – speech or text. Currently, NLP is employed in various ways, from evaluating risks in the social media to answering weather forecast inquiries and more. The spam filter on Gmail, for example, analyzes incoming emails for header information, IP addresses, and content on any spam signals and uses NLP to detect spam.

These are some examples of NLP-powered software which help us in different ways in our daily lives, for example:

- **Personal assistants**: Siri, Cortana, and Google Assistant.
- **Auto-complete**: In search engines (*e.g.* Google, Bing).
- **Spell checking**: Almost everywhere, in your browser, your IDE (*e.g.* Visual Studio), desktop apps (*e.g.:* Microsoft Word).
- **Machine Translation**: Google Translate.

The following includes some of the ways Natural language processing can be used

- Spam detection
- Parts of speech identification
- Sentiment analysis
- Text composition
- Question answering
- Automatic summarization
- Conversational interfaces

## Sentiment Analysis/ Opinion Mining

Sentiment analysis is a common NLP task, which involves classifying texts or parts of texts into a pre-defined sentiment. (*How To Perform Sentiment Analysis in Python 3 Using the Natural Language Toolkit (NLTK) | DigitalOcean*, n.d.). The predefined sentiment could be positive or negative. An example of a sentiment analysis in customer feedback is given in Table 1 below.

**Table 1: Customer feedback and sentiment**

| Customer Feedback Text | Sentiment |
|---|---|
| "This supermarket is great; the staffs are really friendly, and the goods are authentic" | Positive |
| "I would not recommend this supermarket to anyone; their customer service is terrible, and they sell fake stuff" | Negative |

There are three main categories in Sentiment Analysis:

1. Document-level SA: its main objective is, to categorize an attitude text as articulating a positive or negative attitude or sentiment. It deliberates the complete text a basic data unit.
2. Sentence-level SA: its main objective is, to categorize sentiment articulated in individual sentence. The initial stage is to classify either the sentence is subjective nor objective. If the sentence is subjective, Sentence-level sentiment analysis will decide whether the sentence articulates a positive or negative feeling.
3. Aspect-level SA: its main objective is, to categorize the sentiment through feature to the exact features of objects. The primary stage is to classify the objects and their features.

Some of the applications of sentiment analysis can be classified into the following:

- **Business:** Brands use it to build their strategies in the field of marketing to understand the feelings of customers about their goods or the brand, how people respond to their ads and why consumers don't purchase other items.
- **Politics:** It is used in political terms to track political views, detect consistency and inconsistency among statements and actions at government level. It can also be used to forecast outcomes of elections.
- **Public Actions:** Sentiment analysis is used to track and analyze social phenomena, to identify potentially dangerous situations and to assess the blogosphere general mood.

Figure 1: shows the several areas in which sentiment analysis has helped in driving business.

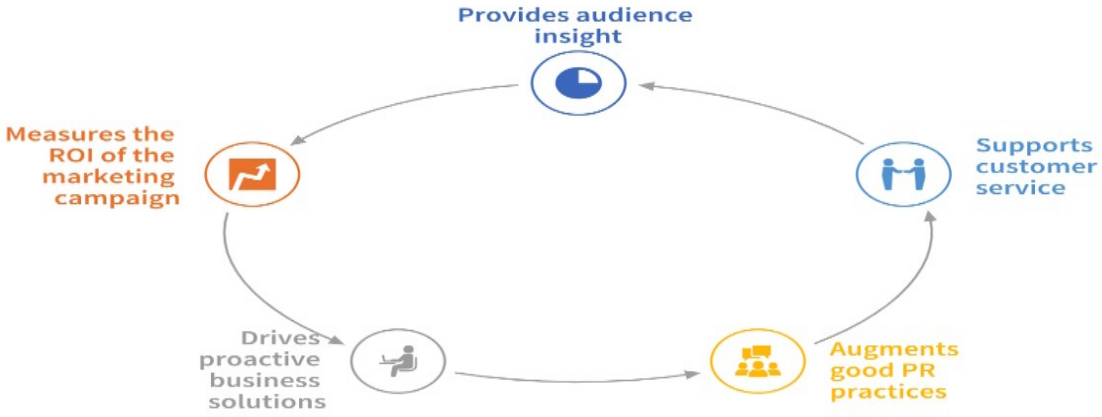


**Figure 1: Advantages of Sentiment Analysis in driving business**
Source: Parul 2019

## Social Media and Sentiment Analysis

Social Media are websites and applications that are designed to allow people to share content quickly, efficiently and in real-time. Social media may not be restricted to apps alone as it actually started with computers, but it is true that most social media users access their tools through apps. Retailers who have leveraged the ability to share photos, opinions, events, etc. in real-time have seen a transformation in the way they do business and those that have integrated a better marketing strategy have seen measurable results. The emergence of Facebook, Twitter, Instagram, blogs, forums and other online networking platforms has created a digital world in which people can socialize and express themselves through a variety of means and applications.(Paltoglou & Thelwall, 2012). Sentiment analysis is a research area which has gained significant attention in recent years and one of the key explanations for this trend is the aforementioned rise in user generated content from social media sites.

For example, Sentiment analysis is widely used in tourism and hospitality to understand the social media comments of travelers. In tourism destination research, sentiment analysis has been proposed as a tool to monitor brands' reputation (Sigala et al., 2012). In this research we make use of Twitter to get opinions of users on a certain topic and automatically analyze those statements to obtain the sentiment polarity.

### Twitter

The use of Twitter as a tool for performing sentiment analysis due to it having a large amount of relevant data, better public opinion, and overall better user interaction, also due to its limit to 120 characters, classification can be done easily and more accurately to give better results. Twitter is a 'microblogging' system that allows you to send and receive short posts called tweets. Tweets can be up to 140 characters long and can include links to relevant websites and resources. Twitter, a social networking site launched in 2006, is undoubtedly one of the most popular social media platforms available today, with over 100 million daily active users and over 500 million tweets sent daily. You can create your own tweets and also retweet what others have tweeted. Twitter have contributed to sentiment analysis in the sense that we can know the real-time opinions of people on a particular subject or hashtag on twitter as it is done in this study and several other research works done with twitter

### Related Works on Sentiment Analysis and Opinion Mining Techniques

Sentiment analysis is a process where the dataset consists of emotions, attitudes or assessment which takes into account the way a human think. Sentiment analysis in the context of computer science with data mining processes is to extract opinions, emotions sentiments in the text such that computers can analyze texts from comments, blogs, review aggregation websites and various types of social media to determine opinions about products and services or domains. The sentiment of the text can be categorized into a set of parameters (such as positive or negative, recommended (thumbs up) or not recommended (thumbs down) or scaled into 1 to 5-star categories. (Shakur, 2013). According to (Gupta et al., 2017),

Sentiment Analysis is a classification technique which derives opinion from the tweets and formulates a sentiment and on the basis of which, sentiment classification is performed. Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model in order to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text pre-processing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data.

Online customer reviews are another significant source of information that is beneficial to both businesses and customers who may want to patronize them in the future. In most cases, the style in which evaluations are provided on webpages is unstructured, making it impossible to manually read through huge numbers of reviews, which is not practical in terms of business. Some Nigerians have performed some predictive analysis using sentiment analysis, for example, (Amusa et al., 2016) submitted the query "Buhari Administration" based on the objective of performing sentiment analysis on drone regulation, and making use of twitter API.

They used the tweets obtained from Twitter to analyze sentiments expressed by Nigerians about the administration of President Muhammadu Buhari. At the end of the analysis, it was discovered that about 45% of the peoples' sentiments were positive towards the policy and programs of the current administration of PMB while about 55% of these sentiments were negative. Results of classification of various sentiments as expressed by Nigerians showed that The NB classifier had overall average prediction accuracy of about 80% with a standard deviation of 0.0258.

The estimated average overall sensitivity and specificity of this classifier were about 83% and 88% respectively which further justified the goodness of the NB model for text mining. Using twitter data set, (Adebisi et al., 2018) attempted to analyze the opinions of Nigerians on some likely presidential candidates (Muhammadu Buhari, AtikuAbubakar, Rabiu Kwankwaso and Ayo Fayose) in the country's 2019 presidential elections. Tweets were manually annotated as positive, negative or neutral by human evaluators for better classification speed and accuracy. The labelled tweets were used to train the Naïve Bayes Classifier which was then used to classify new tweets for the sentiment analysis. The classified twitter data is displayed using pie charts. The results of the analyses showed Buhari had the highest % of tweets over the period. Regarding the 2019 presidential elections, Atiku had the lowest % of negative opinions and the highest % of positive opinions.

### Techniques for Sentiment Analysis

1. Rule-based sentiment analysis: This is also known as Lexicon Based SA. The Rule-based SA uses a dictionary of words labelled by sentiment to determine the sentiment of a sentence. In order to work well, sentiment scores typically need to be combined with additional rules to mitigate sentences containing negations, sarcasm, or dependent clauses.

**Table *1*: below shows a Sentiment Dictionary Example (-1 = Negative, +1 = Positive):**

| Word | Sentiment |
|---|---|
| Good | 0.5 |
| great | 0.8 |
| terrible | -0.8 |
| alright | 0.1 |

2. Machine Learning (ML) based sentiment analysis: In Machine Learning (ML) based SA, we train an ML model to recognize the sentiment based on the words and their order using a sentiment-labelled training set. This approach depends largely on the type of algorithm and the quality of the training data used.

Focusing on building a sentiment analysis classifier using natural language processing. The system enables you to enter sample tweets and returns the polarity. Users can also input a particular hashtag and the system would retrieve a bulk of live tweets from the Twitter API and then the results would be displayed in a bar chart which shows the proportion of tweets in which negative sentiments are expressed to tweets in which positive sentiments are expressed, this diagrammatic representation of data may be useful for individuals, businesses, brands or other organizations that may use the system to get an insight of what people think about a subject as it concerns them.

The following steps needs to be adopted in the process

i. Review of related works in the area of sentiment analysis and opinion mining.
ii. Finding the appropriate methodology, determining the software and hardware requirements and developing a model for the various entities and actions included in the system.
iii. Making use of Natural Language Toolkit (NLTK), a commonly used NLP library in Python, to analyze textual data to prepare a data set for sample tweets from the NLTK package for NLP with different data cleaning methods. Training a model on pre-classified tweets using the Naïve Bayes classification technique to classify the sample tweets into negative and positive sentiments.

iv. Using Tweepy among other tools to search and collect live tweets from twitter and then serving them into the model to produce the sentiment polarity and thereby presenting the result in graphically on the Frontend.

**Error! Reference source not found.** is a diagrammatical representation of some techniques used for sentiment analysis

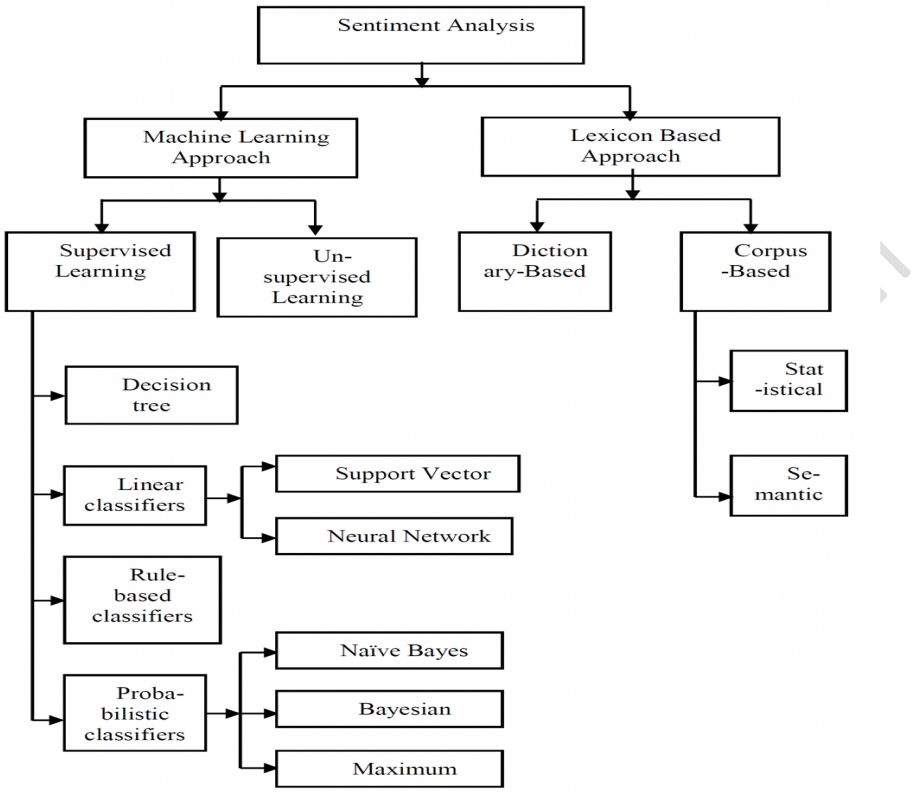


**Figure 2 : Techniques for Sentiment or Opinion Classification (Yousef et al., 2014)**

**Table 2: Table of other Related Works**

| S/N | Authors, year | Title | SA Techniques used | Remarks |
|---|---|---|---|---|
| 1. | Pang and Lee (2010) | A Sentimental Education: Sentiment Analysis using Subjectivity Summarization Based on Minimum Cuts | Various Techniques | Examined the relationship between subjectivity detection and polarity classification |
| 2. | Nathp et al. (2016) | Sentiment Analysis of Twitter Generated Data | Emoticon based analysis | Sentiment Analysis of tweets using emoticons analysis to improve accuracy |
| 3. | Adebisi et al. (2018) | A Simple Opinion Mining Of Some Likely Nigerian Candidates In The 2019 Elections | Naïve Baiyes Classifier | Use SA to deduce politicians that had a higher chance of winning the 2019 elections |
| 4. | Kavya S., & Narasinga R., (2019) | Sentiment Analysis Using Naïve Bayes Classifier | Naïve Bayes | Make use of bag of words concept to develop model that performs SA on twitter data |

| | | | | |
|---|---|---|---|---|
| 5. | Fang and Zhan (2015) | Sentiment Analysis Using Product Review Data | Naïve Bayesian, Random Forest, and Support Vector Machine | Used online reviews from Amazon.com to analyze sentiments. |

To achieve the above, a sentiment analysis model would be built and associate tweets with a positive or a negative sentiment. Dataset is splilt into two parts. The purpose of the first part is to train the model, whereas the other part tests the performance of the model. A simple application was created with HTML & CSS for front end design connected to our model using a python framework known as Flask. The backend would also contain the model and other implementations. The frontend consists of two pages. In the first page, the user can input sample tweet into the text area. The Tweet is then sent to the backend and then the sentiment result is returned from the model and is displayed in the frontend. It also contains a link that takes the user to the next page.

The second page is a chart page to contain a bar chart that shows the sentiment result for tweets obtained from the twitter API through a search term specified by what the user enters into the input field. It consists of a button in which when clicked, it collects the data from twitter, feed it to the model in the backend and then displays the result in form of a bar chart that shows the percentage of negative sentiments over positive sentiments and vice versa.

## Conclusion

The aim of sentiment analysis is to extract opinions, emotions sentiments in the text such that computers can analyze texts from comments, blogs, review aggregation websites and various types of social media to determine opinions about products and services or domains.
After successful completion of this research work, the system developed would be able to achieve the main aim of the work employing machine learning tools and techniques for natural language processing. With the use of a graphical user interface and the diagrammatic representation of data in form of charts on a web server to show sentiment analysis of tweets, the system may be useful for individuals, businesses or brands that may use the system to get insights of what people think about a subject as it concerns them.

## References


1. Adebisi, B. A., Adekanmi, A. A., Uchechukwuand, C. M., & Ojiyovwi, A. F. (2018). A Simple Opinion Mining On Some Likely Nigerian Presidential Candidates in the 2019 Elections. Www.Ijmer.Com |, 8.
2. Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. Journal of Big Data, 2(1). https://doi.org/10.1186/s40537-015-0015-2
3. Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python. International Journal of Computer Applications, 165(9), 29–34. https://doi.org/10.5120/ijca2017914022
4. Nath, A., Sudanp, A., Kumarp, R., & Bhosalep, R. (2016). Sentiment Analysis Of Twitter Generated Data. International Journal of Scientific Engineering and Applied Science (IJSEAS), 2. www.ijseas.com
5. Paltoglou, G., & Thelwall, M. (2012). Twitter , MySpace , Digg : Unsupervised Sentiment Analysis. 3(4). https://doi.org/10.1145/2337542.2337551
6. Sigala, M., Christou, E., & Gretzel, U. (2012). Social media in travel, tourism and hospitality. Theory, Practice and Cases. Surrey, UK: Ashgate Publishing, Ltd. http://docshare02.docshare.tips/files/24579/245792778.pdf
7. Shakur, Z. A. (2013). Opinion Mining/Sentiment Analysis. 40238. African University of Science and technology Research http://repository.aust.edu.ng/xmlui/handle/123456789/397

8. Yousef A., Medhat W., Mohame H., (2014) Sentiment Analysis Algorithms and Applications: A Survey Ain Shams Engineering Journal 5(4) DOI:10.1016/j.asej.2014.04.011

9. Parul, P., (2019) Sentiment Analysis is difficult, but AI may have an answer https://towardsdatascience.com/sentiment-analysis-is-difficult-but-ai-may-have-an-answer-a8c447110357
10. Kavya S., & Narasinga R., (2019) Sentiment Analysis Using Naïve Bayes Classifier International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019 264 Published By: Blue Eyes Intelligence Engineering & Sciences Publication Retrieval Number H6330068819/19©BEIESP