

A Project Report

ON

Sentiment Analysis on Twitter using NLP

SUBMITTED BY-

ANISH PAUL
ADTU/2019-23/BTech(CTIS)/018

HARJIT BARMAN
ADTU/L/2019-23/BTech(CTIS)/027

Under the guidance of:

BHUBNESWAR DAS
ASSISTANT PROFESSOR

Faculty of Engineering & Technology





Faculty of Engineering & Technology

Assam Down Town University

Gandhi Nagar, Panikhati, Guwahati-781026,
Assam

CERTIFICATION FROM GUIDE

This is to certify that the project report entitled “ **Sentiment Analysis on Twitter using NLP** ” submitted by **Anish Paul** bearing Enrollment No. ADTU/2019-23/BTech(CTIS)/018 and **Harjit Barman** bearing Enrollment No. ADTU/L/2019-23/BTech(CTIS)/027 towards the partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering (Cloud Technology & Information Security) under Assam down town University. I hope this project will help in future in practical field. I wish them success in their life.

DATE:

Place: Guwahati

Mr. Bhubneswar Das
Assistant Professor
**Computer Science and
Engineering**
Assam Down Town University
Guwahati 781026



Faculty of Engineering & Technology

Assam Down Town University

Gandhi Nagar, Panikhati, Guwahati-781026,
Assam

CERTIFICATE FROM EXTERNAL EXAMINER

This is to certify that the project report entitled “**Sentiment Analysis on Twitter using NLP and Machine Learning**” submitted by **Anish Paul** bearing Enrollment No. ADTU/2019-23/BTech(CTIS)/018 and **Harjit Barman** bearing Enrollment No. ADTU/2019-23/BTech(CTIS)/027 and towards the partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering (Cloud Technology & Information Security) under Assam down town University is a bonafide research work carried out by him under the supervision and guidance of, **Mr. Bhuneswar Das**, Assistant Professor, Department of Computer Science & Engineering, Assam down town University, Guwahati, has been examined by me and found to be satisfactory. I recommend the thesis for consideration for the award of the degree of Bachelor of Technology in Computer Science and Engineering (Cloud Technology & Information Security) under Assam down town University.

DATE:

.....

Place: Guwahati

External Examiner

.....

Internal Examiner

SELF DECLARATION

We, **ANISH Paul** bearing Enrollment No. ADTU/2019-23/BTech(CTIS)/018 and **HARJIT BARMAN** bearing Enrollment No. ADTU/L/2019-23/BTech(CTIS)027 hereby declare that the present thesis entitled “ **Sentiment Analysis using NLP**” is an work carried out in the Department of Computer Science and Engineering, Assam Down Town University, Guwahati with exception of guidance and suggestions received from our supervisor, **Mr. Bhubneswar Das**, Assistance Professor, Department of Computer Science & Engineering, Assam Down Town University, Guwahati. The data and the findings discussed in the thesis are the outcome of our research work. This project report is being submitted to Assam Down Town University for the degree of **Bachelor of Technology in Computer Science & Engineering (Cloud Technology & Information Security)**.

Anish Paul

Enrollment No.: ADTU/2019-23/BTech(CTIS)/018

Harjit Barman

Enrollment No.: ADTU/L/2019-23/BTech(CTIS)/027

ABSTRACT

Twitter sentimental analysis is a beneficial aspect in many industrial applications. As it's help to know what's going on twitter .what kind of tweets people are posting in twitter. Sentiment analysis is the process of determining the emotional tone behind a piece of text, particularly in the context of social media such as Twitter. It involves using natural language processing (NLP) techniques to identify and extract subjective information from text data. One common approach to sentiment analysis on Twitter is to use machine learning algorithms to classify tweets as positive, negative, or neutral based on the words and phrases used. This can be done by training a classifier on a large dataset of labeled tweets, and then using the trained classifier to predict the sentiment of new tweets. There are various challenges to performing sentiment analysis on Twitter, including the use of slang, abbreviations, and emoticons, as well as the presence of irony and sarcasm. To address these challenges, researchers have developed techniques such as sentiment lexicons, which are lists of words and phrases with associated sentiment scores, and advanced machine learning algorithms that can handle contextual information. Overall, sentiment analysis on Twitter can be used to gain insights into public opinion about a particular topic or product, as well as to track and respond to customer sentiment in real world.

CHAPTERS

1. Introduction

- 1.1.** Overview of the project
- 1.2.** Motivation
- 1.3.** Scope & Objective
- 1.4.** Existing Approaches
- 1.5.** Problem Definition
- 1.6.** Proposed Approaches

2. Project Analysis

- 2.1.** Project Requirement Analysis
- 2.2.** Advantage & Disadvantages
- 2.3.** Project feasibility Study

3. Literature Review

- 3.1.** Existing Methodologies
- 3.2.** Related Work

4. Project Implementation

- 4.1.** Description of the Software used

5. Result Analysis

- 5.1.** Experimental setup
- 5.2.** Performance comparison with existing system

6. Conclusion & Future Scope

References

1. INTRODUCTION

1.1. Overview of the project

Natural Language Processing (NLP) is a unique subset of Machine Learning which cares about the real life unstructured data. Although computers cannot identify and process the string inputs, the libraries like Pandas, Seaborn, String NLTK, Matplotlib, Numpy and many others found a way to process string mathematically.

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text ,speech, tweets, and database source through NLP it is also known as opinion mining, it determines whether a piece of writing is positive, negative or neutral. Today most of the people use social networking sites to express their opinion about something .companies have been receiving polls about the products they manufacture. the sentiment analysis is done using various machine learning technique.

Twitter is a platform where most of the people express their feelings towards the current context. Sentiment analysis is an NLP technique that allows us to classify if a text, tweet or comment is either positive, neutral or negative. As humans, we can guess the sentiment of a sentence whether it is positive or negative. Similarly, in this project we are going to show you how to train and develop a simple Twitter Sentiment Analysis using python and NLP libraries.

1.2. Motivation

The motivation of Sentiment Analysis is to harness this data in order to obtain important information regarding public opinion, that would help make smarter business decisions, political campaigns and better product consumption.

Tweets sometimes express opinions about different topics. These opinions are important in many business-related decisions and even political sentiments about a candidate.

Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. E.g. Election Polls

Organizations can also use this to gather critical feedback about problems in newly released products. E.g. Brand Management

1.3. Scope & Objective

Sentiment analysis looks at the emotion expressed in a text. It is commonly used to analyze customer feedback, survey responses, and product reviews. Social media monitoring, reputation management, and customer experience are just a few areas that can benefit from sentiment analysis.

The objective of the project is to analyze the tweets and classify whether the tweet is positive or negative sentiment. The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets.

Formally, given a training sample of tweets and labels, where label '1' denotes the tweet is racist/sexist and label '0' denotes the tweet is not racist/sexist, your objective is to predict the labels on the test dataset.

For training the models, we provide a labelled dataset of 31,962 tweets. The dataset is provided in the form of a csv file with each line storing a tweet id, its label and the tweet.

In this analysis we're going to process text based data, machines can't understand text-oriented data so we'll convert the text to vectors and proceed further.

1.4. Existing Approaches

1.4.1. Transforming Dataset using TF-IDF Vectorizer

Scikit-learn's `TfidfTransformer` and `TfidfVectorizer` aim to do the same thing, which is to convert a collection of raw documents to a matrix of TF-IDF features. The differences between the two modules can be quite confusing and it's hard to know when to use which. This article shows you how to correctly use each module, the differences between the two and some guidelines on what to use when.

1.4.2. Initialize CountVectorizer

In order to start using `TfidfTransformer` you will first have to create a `CountVectorizer` to count the number of words (term frequency), limit your vocabulary size, apply stop words and etc. The code below does just that.

1.4.3. Bag of Words

The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.

1.4.4. Test Data Transformation

Performing data transformations is a bit complex, as it cannot be achieved by writing a single SQL query and then comparing the output with the target. For ETL Testing Data Transformation, you may have to write multiple SQL queries for each row to verify the transformation rules.

To start with, make sure the source data is sufficient to test all the transformation rules. The key to perform a successful ETL testing for data transformations is to pick the correct and sufficient sample data from the source system to apply the transformation rules.

1.5. Problem Defination

Sentiment analysis is fascinating for real-world scenarios. However, it faces many problems and challenges during its implementation.

It is tough if compared with topical classification with a bag of words features performed well.

In many cases, words or phrases express different meanings in different contexts and domain.

The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it.

So, the task is to classify racist or sexist tweets from other tweets. Formally, given a training sample of tweets and labels, where label ‘1’ denotes the tweet is racist/sexist and label ‘0’ denotes the tweet is not racist/sexist, your objective is to predict the labels on the given test dataset.

1.6. Proposed Approach

1.6.1. Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

1.6.2. Data Visualization

Data visualization is a technique that uses an array of static and interactive visuals within a specific context to help people understand and make sense of large amounts of data. The data is often displayed in a story format that visualizes patterns, trends and correlations that may otherwise go unnoticed. Data visualization is regularly used as an avenue to monetize data as a product. An example of using monetization and data visualization is Uber. The app combines visualization with real-time data, so customers can request a ride

Visualization lets you comprehend vast amounts of data at a glance and in a better way. It helps to understand the data better to measure its impact on the business and communicates the insight visually to internal and external audiences. Decisions can't be made in a vacuum.

1.6.3. Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it.

1.6.4. Model Training

A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is used to modify the model.

This iterative process is called “model fitting”. The accuracy of the training dataset or the validation dataset is critical for the precision of the model.

Model training in machine language is the process of feeding an ML algorithm with data to help identify and learn good values for all attributes involved. There are several types of machine learning models, of which the most common ones are supervised and unsupervised learning.

2. Project Analysis

2.1. Project Requirement Analysis

2.1.1. Software Requirement

Operating System	Windows 10/11
Programming language	Python 3.10
Code IDE	Anaconda/Jupyter Notebook

2.1.2. Required Python Libraries

- ◆ Numpy - To perform a wide variety of mathematical operation.
- ◆ Matplotlib - Is used for data visualization and graphical plotting.
- ◆ Seaborn - Build on matplotlib with similar functionality.
- ◆ Re - Is used as regular expression to find particular pattern and process it.
- ◆ Nltk- A natural language processing toolkit module associated in anaconda.
- ◆ Panda - To perform data manipulation and analysis

2.1.3. Hardware Requirement

Processor	Intel I5 8 th gen
Processor Speed	1.60 GHz
Hard Disk	1TB SSD
RAM	8GB

2.2. Advantages & Disadvantages

Advantages of sentiment analysis:

- Accurate, unbiased results
- Enhanced insights
- More time and energy available for staff do to higher-level tasks
- Consistent measures you can use to track sentiment over time

Disadvantages of sentiment analysis:

- Best for large and numerous data sets. To get real value out of sentiment analysis tools, you need to be analyzing large quantities of textual data on a regular basis.
- Sentiment analysis is still a developing field, and the results are not always perfect. You may still need to sense-check and manually correct results occasionally.

2.3. Feasibility Study

2.3.1. Operational Feasibility

In Operational Feasibility degree of providing service to requirements is analyzed along with how much easy product will be to operate and maintenance after deployment. Along with this other operational scopes are determining usability of product, Determining suggested solution by software development team is acceptable or not etc.

2.3.2. Technical Feasibility

In Technical Feasibility current resources both hardware software along with required technology are analyzed/assessed to develop project. This technical feasibility study gives report whether there exists correct required resources and technologies which will be used for project development. Along with this, feasibility study also analyzes technical skills and capabilities of technical team, existing technology can be used or not, maintenance and up-gradation is easy or not for chosen technology etc.

We are going to use the following software's to develop our system –

- 1.Python 3.10
- 2.NLTK
- 3.Mat-plot-lib
- 4.Regular expression
- 5.Pandas
- 6.Sea-born
- 7.Numpy
- 8.Anaconda
- 9.Jupyter Notebook
- 10.Windows 10/11 OS

2.3.3. Economic Feasibility

This is a crucial factor to take into account when creating a project. We choose the technology based on how affordable it is. Since the initiative is intended to meet the demands of the users, it is not financially problematic. The project only needs the barest minimum of resources. It doesn't require any more resources. who can then operate it.

3. Literature Review

3.1. Existing Methodology

3.1.1. Machine Learning Approach

Machine learning approach uses machine learning algorithm and it is classified as supervised learning and unsupervised learning. Supervised learning requires desired output to compare with the actual output. While unsupervised learning does not require any desired output rather it uses previous experience and data to improve its accuracy.

It uses different types of algorithm to carry out the sentiment analysis. It includes training the particular portion of dataset and then using the remaining portion of dataset to test for the result. Majorly used algorithm is:

- Naive Bayes

Naive Bayes algorithm is derived from Bayes' theorem. It consists of a family of algorithms. Bayes' theorem computes the probability of given set using already calculated probabilities.

The diagram shows the formula for Bayes' theorem: $P(c|x) = \frac{P(x|c)P(c)}{P(x)}$. Arrows point from labels to the components of the formula: 'Likelihood' points to $P(x|c)$, 'Class Prior Probability' points to $P(c)$, 'Posterior Probability' points to $P(c|x)$, and 'Predictor Prior Probability' points to $P(x)$.

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

3.1.2. Lexicon Based Approach

Lexicon based approach, Machine learning approach and Hybrid approach are the classification of Sentiment Analysis. Lexicon based approach divides the entire document into lexemes which is used to examine the sentences. Lexicon based approach is further classified as corpus based approach and dictionary based approach. Corpus based approach finds out the polarity of the sentence as negative, positive and neutral. Positive – beautiful, best, excellent, etc. and Negative – bad, disgusting, irritating, etc.

Dictionary based approach is a mathematical approach for measuring the feeling that the sentences conveys to the reader.

As mentioned above, lexicon based approach basically deals with lexeme i.e. tokens or words. It splits the sentence into tokens and processes them. These words are classified as positive or negative opinions.

Classification of Lexicon based approach is as follows:

- **Corpus based approach**

It came into the picture to resolve the problems of dictionary based approach. It is less efficient than dictionary based approach because there is a need to make a large corpus for covering English words which is a difficult task.

- **Dictionary based approach**

It provides better efficiency than corpus based approach. It uses a dictionary which consists of all the synonyms and antonyms of each words. It cannot find the opinion with domain specific orientation

3.2. Related Work

Till now, many researchers have worked on the analysis of product reviews. In paper [3] Elli, Maria and Yi-Fan extracted sentiments from the reviews and analyzed the result to build a business model. They claimed that the proposed tools gives high enough accuracy due to its robustness. They made their decision more accurate by using business analytics. Other notable work includes emotion detection, gender based on the names and fake review detection. The commonly used programming languages are Python and R. Monomial Naïve Bayes along with Support Vector Machine were used. In paper [4] supervised learning algorithms are used to generate ratings on a numerical scale with the help of text only. Their total dataset is divided into 70% training data and 30% testing data. They used various classifiers for their module. The author in Paper [5] continues the idea of Sentiment Analysis and Natural Language Processing. Reviews are labeled as positive or negative using Naïve Bayes' and decision list classifier. The author in paper [6] builds a system that helps to visualise the sentiment in the form of charts. Scraping was used to collect usable data from product's URL. Author generated sentiments using Naive Bayes and Support Vector Machine algorithms. The paper does not mention accuracy anywhere and the output is represented in statistical form. In the paper [7] writer predicts product rating based on concept of bag-of words. These models utilizes unigrams and bigrams. The Amazon video game reviews subset of UCSD Time Based model didn't work well, since the variance in the average rating between each day, month or year was relatively small. Unigram produce better results as compared to Bigram. Unigram results had an approximately 16.00% better performance than bigrams.

4. Project Implementation

4.1. Description of the Software used

4.1.1. Anaconda Navigator

Anaconda is an open-source distribution of the Python and R programming languages for data science that aims to simplify package management and deployment. Package versions in Anaconda are managed by the package management system, conda, which analyzes the current environment before executing an installation to avoid disrupting other frameworks and packages.

The Anaconda distribution comes with over 250 packages automatically installed. Over 7500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI (graphical user interface), Anaconda Navigator, as a graphical alternative to the command line interface. Anaconda Navigator is included in the Anaconda distribution, and allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages, install them in an environment, run the packages and update them.

4.1.2. Jupiter Notebook

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

5. Result Analysis

5.1 Experimental setup

5.1.1. Loading the dataset

Data loading is the process of copying and loading data or data sets from a source file, folder or application to a database or similar application. It is usually implemented by copying digital data from a source and pasting or loading the data to a data storage or processing utility.

```
In [2]: df = pd.read_csv('Twitter Sentiments.csv')
df.head()

Out[2]:
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation

5.1.2. Preprocessing the dataset

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Remove twitter handles (@user)

	id	label	tweet	clean_tweet
0	1	0	@user when a father is dysfunctional and is s...	when a father is dysfunctional and is so sel...
1	2	0	@user @user thanks for #lyft credit i can't us...	thanks for #lyft credit i can't use cause th...
2	3	0	bihday your majesty	bihday your majesty
3	4	0	#model i love u take with u all the time in ...	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation	factsguide: society now #motivation

Remove special characters, numbers and punctuations

	id	label	tweet	clean_tweet
0	1	0	@user when a father is dysfunctional and is so sel...	when a father is dysfunctional and is so sel...
1	2	0	@user @user thanks for #lyft credit i can't us...	thanks for #lyft credit i can t use cause th...
2	3	0	bihday your majesty	bihday your majesty
3	4	0	#model i love u take with u all the time in ...	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation	factsguide society now #motivation

5.1.3. Exploratory Data Analysis

Exploratory data analysis (EDA) involves using graphics and visualizations to explore and analyze a data set. The goal is to explore, investigate and learn, as opposed to confirming statistical hypotheses.

Exploratory data analysis is a powerful way to explore a data set. Even when your goal is to perform planned analyses, EDA can be used for data cleaning, for subgroup analyses or simply for understanding your data better. An important initial step in any data analysis is to plot the data.

Visualize the frequent words



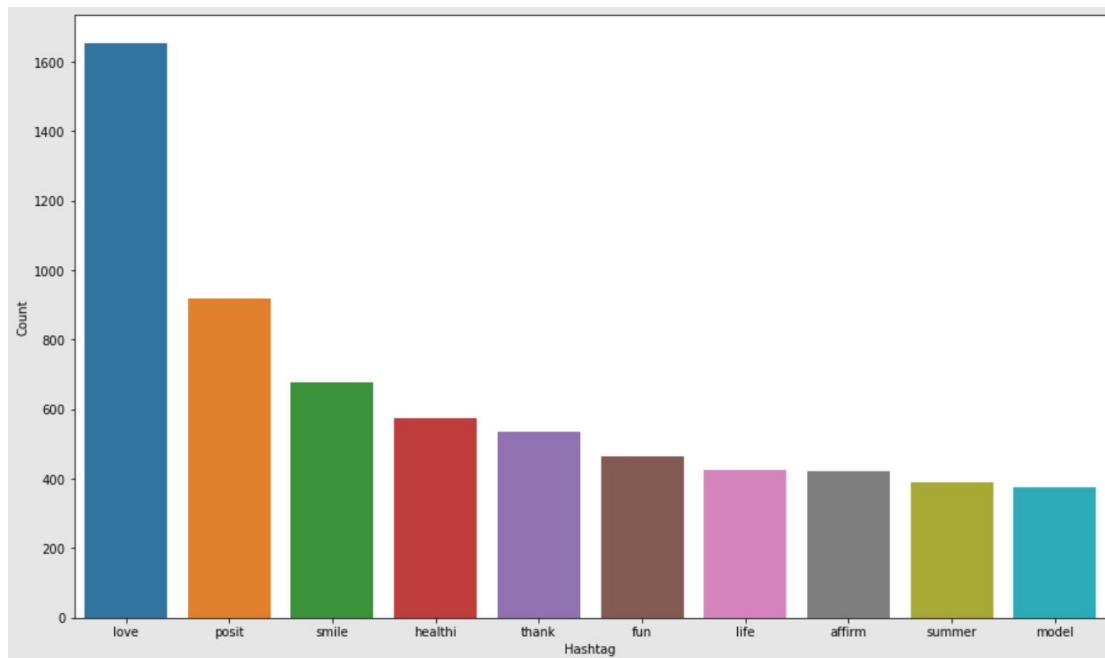
5.2. Performance comparision with existing system

5.2.1 Extracting positive words

List of positive hashtag

	Hashtag	Count
0	run	72
1	lyft	2
2	disapoint	1
3	getthank	2
4	model	375

Graph of positive hashtag

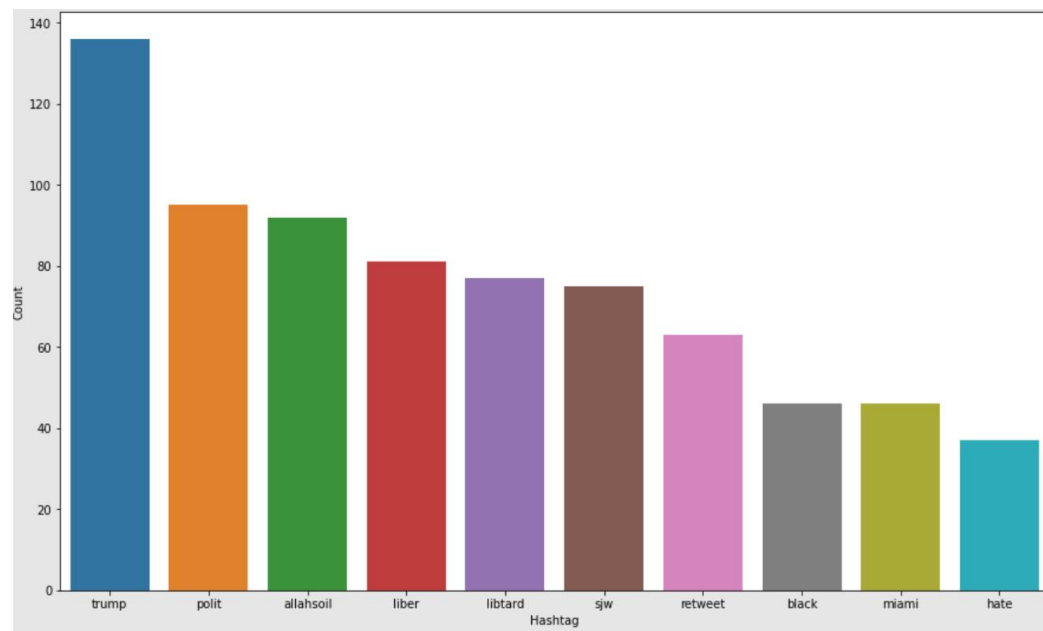


5.2.2. Extracting negative hashtage

List of negative hashtag

	Hashtag	Count
0	cnn	10
1	michigan	2
2	tcot	14
3	australia	6
4	opkillingbay	5

Graph of positive hashtag



5.2.2. Model Model Training

```
In [39]: # testing
pred = model.predict(x_test)
f1_score(y_test, pred)
```

```
Out[39]: 0.49763033175355453
```

```
In [42]: accuracy_score(y_test, pred)
```

```
Out[42]: 0.9469403078463271
```

```
In [43]: # use probability to get output
pred_prob = model.predict_proba(x_test)
pred = pred_prob[:, 1] >= 0.3
pred = pred.astype(np.int)

f1_score(y_test, pred)
```

```
Out[43]: 0.5545722713864307
```

```
In [44]: accuracy_score(y_test, pred)
```

```
Out[44]: 0.9433112251282693
```

```
In [47]: pred_prob[0][1] >= 0.3
```

```
Out[47]: False
```

6. Conclusion & Future Scope

Conclusion

Twitter sentiment analysis is the process of using natural language processing and machine learning techniques to identify and extract subjective information from social media posts, particularly tweets. It involves classifying tweets as positive, negative, or neutral based on the words and phrases used in the tweets.

There are a number of approaches to Twitter sentiment analysis, including rule-based approaches, which rely on pre-defined rules to classify tweets, and machine learning-based approaches, which use algorithms to learn patterns in the data and make predictions.

One of the main challenges of Twitter sentiment analysis is the use of slang, abbreviations, and emoticons, which can make it difficult for algorithms to accurately interpret the sentiment of a tweet. Another challenge is the lack of context, as tweets are often short and may not provide enough information to accurately gauge the sentiment of the tweet.

Overall, Twitter sentiment analysis can be a useful tool for understanding the public sentiment towards a particular topic or brand, but it is important to be aware of the limitations and challenges of this type of analysis.

Future Scope

In this project, for the time being we are able to extract positive and negative words. In future we plan to identify the whole tweet as positive and negative. And use this for better positive future.

And there is a wide range of potential for Twitter Sentiment Analysis in future -

- Marketing and Advertising
- Political Analysis
- Customer Service
- Financial Analysis
- Social Listening

References

- [1] Diana Maynard, Adam Funk. Automatic detection of political opinions in tweets. In: Proceedings of the 8th international conference on the semantic web, ESWC'11; 2011.
- [2] (2015) Machine Learning with Naïve Bayes Classifier [Online]. Available: <http://blog.datumbox.com/machine-learning/tutorial-the-naive-bayes-text-classifier/>
- [3] Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016
- [4] Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp,,s Ratings Based on Text Reviews." (2015).
- [5] Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning." Swarthmore College (2013).
- [6] Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110.
- [7] Chen, Weikang, Chihhung Lin, and Yi-Shu Tai. "Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015).
- [8] Shaikh, Tahura, and DeepaDeshpande. "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.", (2016)
- [9] Nasr, Mona Mohamed, Essam Mohamed Shaaban, and Ahmed Mostafa Hafez. "Building Sentiment analysis Model using Graphlab." IJSER, 2017
- [10] Text mining for yelp dataset challenge; Mingshan Wang; University of California San Diego, (2017)
- [11] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Spring